

Basis Adaptive Sample Efficient Polynomial Chaos (BASE-PC)

Jerrad Hampton, Alireza Doostan*

*Aerospace Engineering Sciences Department, University of Colorado, Boulder, CO
80309, USA*

Abstract

For a large class of orthogonal basis functions, there has been a recent identification of expansion methods for computing accurate, stable approximations of a quantity of interest. This paper presents, within the context of uncertainty quantification, a practical implementation using basis adaptation, and coherence motivated sampling, which under assumptions has satisfying guarantees. This implementation is referred to as Basis Adaptive Sample Efficient Polynomial Chaos (BASE-PC). A key component of this is the use of anisotropic polynomial order which admits evolving global bases for approximation in an efficient manner, leading to consistently stable approximation for a practical class of smooth functionals. This fully adaptive, non-intrusive method, requires no *a priori* information of the solution, and has satisfying theoretical guarantees of recovery. A key contribution to stability is the use of a presented correction sampling for coherence-optimal sampling in order to improve stability and accuracy within the adaptive basis scheme. Theoretically, the method may dramatically reduce the impact of dimensionality in function approximation, and numerically the method is demonstrated to perform well on problems with dimension up to 1000.

Keywords: Polynomial Chaos, Orthogonal Polynomials, Uncertainty Quantification, Compressive Sensing, Basis Adaptation, Importance Sampling

*Corresponding Author: Alireza Doostan

Email address: `alireza.doostan@colorado.edu` (Alireza Doostan)

1. Introduction

A reliable approach to analyzing complex engineering systems requires understanding how various Quantities of Interest (QoI) depend upon system inputs that are often uncertain; where a poor understanding will lead to poor executive decisions. Uncertainty Quantification (UQ) [1, 2, 3] is a field that aims at addressing these issues in a practical and rigorous manner, giving a meaningful characterization of uncertainties from the available information and admitting efficient propagation of these uncertainties for a quantitative validation of model predictions.

Probability is a natural framework for modeling uncertainty, wherein we assume uncertain inputs are represented by a d -dimensional random vector $\Xi := (\Xi_1, \dots, \Xi_d)$ with some joint probability density function $f(\xi)$ supported on Ω , where we further assume that the coordinates of Ξ are independent. In this manner, the scalar QoI to be approximated, here denoted by $u(\Xi)$, is modeled as a fixed but unknown function of the input. In this work we approximate $u(\Xi)$, assumed to have finite variance, by a spectral expansion in multivariate basis functions, each of which is denoted by $\psi_k(\Xi)$, and are naturally chosen to be orthogonal with respect to the distribution of Ξ [4, 5]. We focus here on the case that ψ_k are polynomials, a method referred to as a Polynomial Chaos (PC) expansion [1, 4],

$$u(\Xi) = \sum_{k=0}^{\infty} c_k \psi_k(\Xi). \quad (1)$$

We note that the independence assumption for the coordinates of Ξ may be removed if care is taken in prescribing orthogonal basis functions ψ_k , although we do not consider any such examples here.

For computation, we allow an arbitrary number of input dimensions d but assume u can be accurately approximated in some relatively small set of basis functions. Let $\mathbf{k} = (k_1, \dots, k_d)$ be a vector such that $k_i \in \mathbb{N} \cup \{0\}$ represents the order of the polynomial $\psi_{k_i}(\Xi_i)$, which is orthonormal with respect to the distribution of Ξ_i . For instance, when Ξ_i follows a uniform or Gaussian distribution, $\psi_{k_i}(\Xi_i)$ are normalized Legendre or Hermite polynomials, respectively [4, 5]. For a d -dimensional vector \mathbf{k} , the d -dimensional polynomial $\psi_{\mathbf{k}}(\Xi)$ is then constructed by the tensorization of $\psi_{k_i}(\Xi_i)$, where

k_i is the i th coordinate of \mathbf{k} . Specifically,

$$\psi_{\mathbf{k}}(\Xi) = \prod_{i=1}^d \psi_{k_i}(\Xi_i).$$

In this work we select basis functions in a manner that iteratively adjusts parameters that define a basis. Specifically, we consider a definition of anisotropic total order [6] using one parameter, p_i , per dimension. We combine these into a vector, $\mathbf{p} := (p_1, \dots, p_d)$, so that an order- \mathbf{p} basis is defined by a related set of $\mathbf{k} = (k_1, \dots, k_d)$, specifically

$$\mathcal{B}_{\mathbf{p}} := \left\{ \psi_{\mathbf{k}} \left| \sum_{i=1}^d \frac{k_i}{p_i} \leq 1 \right. \right\}. \quad (2)$$

This basis definition has a number of parameters that scales with dimension, and which we will repeatedly modify to improve the quality of our polynomial approximation. We note that if all $p_i = p$, then the order- \mathbf{p} basis is identical to a total order basis of order p . We also note that this basis can have an additional hyperbolicity parameter associated with it as considered in [7], although we do not consider any such parameter here. Heuristically, we expect most p_i to be low and only a few to be relatively high, allowing a basis that faithfully approximates the QoI with relatively few basis functions compared to a total order basis with an order that is able to achieve the same accuracy in the reconstruction. Often, the subscript on \mathcal{B} is omitted; replaced with a scalar index related to iterative adjustment; or replaced with a bound on approximation error achieved in that basis; and this should not be confusing in context. For the remainder of this text, we refer to an order- \mathbf{p} basis as an anisotropic order basis.

We use $|\mathcal{B}|$ to denote the total number of basis functions in a set \mathcal{B} , indexed in an arbitrary manner for $k = \{1, \dots, |\mathcal{B}|\}$, while the vector \mathbf{k} specifically identifies the basis function by determining the order in each dimension. This facilitates a polynomial surrogate approximation to u for any basis set \mathcal{B} , given by

$$u(\Xi) \approx \sum_{k=1}^{|\mathcal{B}|} c_k \psi_{\mathbf{k}}(\Xi). \quad (3)$$

The error introduced by this truncation is referred to as *truncation error*, and converges to zero – in the mean squares sense as basis functions are added –

when

$$c_k = \mathbb{E}(u(\Xi)\psi_k(\Xi)). \quad (4)$$

Here, \mathbb{E} denotes the mathematical expectation operator. Without any *a priori* information as to what \mathcal{B} should be, we seek to identify \mathcal{B} based solely on solution characteristics as revealed by computed coefficients, $\{c_k\}$.

Identifying an optimal \mathcal{B} first involves identifying a scalar quantity to optimize. In the present work, this quantity is related to a cross-validated error computed via ℓ_1 -minimization using non-intrusive methodology [8, 9]. Specifically, for a fixed basis, to identify the PC coefficients $\mathbf{c} = (c_1, \dots, c_{|\mathcal{B}|})^T$ in (3) we consider a sampling-based method. This method does not require changes to deterministic solvers for u as we generate realizations of Ξ to identify $u(\Xi)$, or perform a related importance sampling as in [10, 11]. We denote the i th such realizations as $\boldsymbol{\xi}^{(i)}$ and $u(\boldsymbol{\xi}^{(i)})$, respectively. We let N denote the number of samples of the QoI which we utilize, and define,

$$\mathbf{u} := (u(\boldsymbol{\xi}^{(1)}), \dots, u(\boldsymbol{\xi}^{(N)}))^T; \quad (5)$$

$$\Psi(i, j) := \psi_j(\boldsymbol{\xi}^{(i)}), \quad (6)$$

where we refer to Ψ as the *measurement matrix* associated with \mathcal{B} . These definitions imply the matrix equality $\Psi\mathbf{c} = \mathbf{u}$, or more generally that this equality holds approximately. We also introduce a diagonal positive-definite matrix \mathbf{W} such that $\mathbf{W}(i, i) = w(\boldsymbol{\xi}^{(i)})$, a function of $\boldsymbol{\xi}^{(i)}$, is determined by our sampling strategy in the manner of basis-dependent importance sampling; see [10, 11] and Section 2.2. This weighting and the corresponding importance sampling are described in Section 2.2. Here we employ compressive sampling, specifically the Basis Pursuit Denoising [12, 13, 14, 15] interpretation of ℓ_1 -minimization, to compute $\hat{\mathbf{c}}$, our identified coefficients,

$$\hat{\mathbf{c}} := \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{c}\|_1 \text{ subject to } \|\mathbf{W}(\mathbf{u} - \Psi\mathbf{c})\|_2 \leq \delta\|\mathbf{W}\mathbf{u}\|_2, \quad (7)$$

where δ is set via cross-validation [9]. The optimization in (7) may be solved efficiently via interior point methods, where we utilize here an implementation of SPGL1 [16] that is slightly modified for repeated utilization, as our method depends on repeatedly computing these coefficients for various bases. We refer to $\mathbf{W}\Psi$, as the *design matrix*, denoted by \mathbf{D} , i.e.

$$\mathbf{D} := \mathbf{W}\Psi. \quad (8)$$

We note that we use (7) here to compute coefficients, and that this is motivated from the robustness of compressive sensing wherein the number of samples is small compared to the number of basis functions. That is solutions to (7) are robust to including unnecessary basis functions, defined as basis functions whose inclusion does not significantly reduce the error of the reconstructed surrogate. This is important, as though we seek to limit the number of unnecessary basis functions, computing solutions via (7) insures that having unnecessary basis functions has a relatively small effect on the number of samples needed to compute an accurate surrogate. We note that our method to identify this basis generally reduces the number of basis functions considerably, potentially to the point where the number of basis functions is exceeded by the number of samples. In this sense, the method presented here is not clearly interpreted in terms of compressive sensing, although the theoretical guarantees with regards to sparsity concerning solutions computed via ℓ_1 -minimizations still apply.

Recalling (3), we denote our surrogate approximation to u in terms of these computed coefficients, $\{\hat{c}_k\}$, by

$$\hat{u}(\Xi) := \sum_{k=1}^{|\mathcal{B}|} \hat{c}_k \psi_k(\Xi). \quad (9)$$

Here the surrogate reconstruction of u , denoted \hat{u} , is computed iteratively via solution to (7) repeated over different potential reconstruction bases and available samples. We measure our error by relative root-mean-square error (RRMSE), defined by

$$\text{RRMSE}(\hat{u}) := \frac{\sqrt{\mathbb{E}(\hat{u}(\Xi) - u(\Xi))^2}}{\sqrt{\mathbb{E}(u^2(\Xi))}}. \quad (10)$$

Our identification of a basis is done so as to minimize a validated estimate of $\text{RRMSE}(\hat{u})$, i.e. we select a basis that with its corresponding computed coefficients returns the lowest estimate of $\text{RRMSE}(\hat{u})$ from the set of considered bases. This estimate is computed from repeated solution of (7) for different subsamples of our total pool of available samples. The class of potential anisotropic order bases depends on the computed coefficients, $\{\hat{c}_k\}$, as well as the dimension and order of the associated basis functions. Heuristically, p_k is increased in dimensions with basis functions having high order in that dimension and large magnitude solution coefficients. Conversely, p_k

is decreased in dimensions where basis functions having high order in that dimension are associated with solution coefficients having low magnitude.

This is a heuristic similar to that utilized in [17], and typically favors dimensions with more local variance as in the approach of [18]. From this coefficient magnitude information, the basis is adapted from a basis denoted \mathcal{B}_0 to one denoted \mathcal{B}_1 . If specified, this basis adaptation also includes an increase to the dimension of the PC basis. During this step, several potential \mathcal{B}_1 are generated and tested. From this set of potential bases the basis giving the lowest cross-validated approximation error is kept. With this error minimizing basis, new samples are identified that assure a low coherence for the aggregate samples with respect to this basis as in [10, 11]. With these additional samples, the basis may then be updated again, and the process of basis adaptation and sample identification may be repeated in an iterative manner.

Recalling c_k from (4), we assume the error model

$$\begin{aligned} u(\Xi) &= \sum_{k=1}^{|\mathcal{B}|} c_k \psi_k(\Xi) + \epsilon(\Xi), \\ &\approx \sum_{k=1}^{|\mathcal{B}|} \hat{c}_k \psi_k(\Xi) + \epsilon(\Xi), \\ &= \hat{u}(\Xi) + \epsilon(\Xi), \end{aligned} \tag{11}$$

noting that the robustness of solutions with regards to model or measurement errors has been investigated [19, 20, 10, 11]. Generally, we seek to guarantee that $\text{RRMSE}(\hat{u})$ is close to $\sqrt{\mathbb{E}(\epsilon^2(\Xi))}/\sqrt{\mathbb{E}(u^2(\Xi))}$. As the number of basis functions used for our approximation increases, the error arising from performing regression with an incomplete set of basis functions is shown for examples to converge to zero more rapidly than for comparable non-adaptive bases, both in terms of the number of samples needed to compute the approximation, and with regards to the number of basis functions used in the approximation.

In summary, to achieve any specified approximation error, the design and measurement matrices for the basis adaptive approach require significantly fewer entries than the corresponding non-adaptive approach. These methods are referred to collectively as Basis Adaptive Sample Efficient Polynomial Chaos (BASE-PC), and are presented in detail in Section 2. While Compressive Sensing [21, 14, 22, 23] can handle a relatively large set of basis

functions using the sparsity promoted in solutions to (7), and do so within the context of UQ [8, 9, 24, 25, 26, 27, 28, 29, 30, 31], the number of basis functions is still responsible for algorithmic bottlenecks, and a reduction of $|\mathcal{B}|$ through the shaping of the operative basis can produce significant gains in accuracy [18, 24, 32, 7, 17].

Though not considered here, as in [33, 34], an independent column weighting, \mathbf{V} , may be used to reduce the contribution of higher order polynomials and give a more stable approximation for high order models, particularly if interpolation is desired in place of the regression considered here. Further, the results of [33] may assist with identifying appropriate ratios of samples to basis functions for stable, alias-free approximations in such cases. We also note that the inclusion of derivative information as in [35] falls within the coherence and coherence-optimal sampling framework, although we do not consider any examples that utilize derivative information here. Noting that a truncation to a finite-dimensional problem is necessary for computation, d may be infinite within similar contexts as in [36], although we assume in this work that some truncation to a finite dimension d is identified before computation is performed. The infinite dimensional results and framework of [36] also directly corresponds to our use of ℓ_1 -minimization on subsets of the infinite set of basis functions which exists in the context of polynomial approximation, even when d is finite.

1.1. Contributions of This Work

This work combines and advances several results from recent developments in PC into a single practical implementation designed to promote stability and convergence with theoretical guarantees. As an extension of previous related work, the main contributions of this study are as follows.

The sampling distributions in [10, 11] are given expanded utility to the practical case where the reconstruction basis may change. This is done by identifying a novel correction sampling that retains all previously generated samples, while giving aggregate sample pools from an appropriate distribution that guarantees a stability in the approximations, i.e. that allows an adaptation of the sampling distribution to similarly adapting bases. This proposed use of correction sampling within importance sampling is novel to the authors' knowledge.

This method also provides an approach to adaptive PC that builds upon and differs conceptually from adaptations in the stochastic space [37, 18, 38], and utilizes a different approach to basis adaptivity when compared to other

proposed methods for adapting the basis [17, 24, 32, 39, 7]. Key to this adaptation is the use of anisotropic total order, which is described by d parameters, allowing for an efficient approach to adaptation, while being robust with regards to the functions it is capable of approximating. Specifically, it uses a global basis that is a specific version of those considered in [7], while using different methods for sampling and basis identification. This basis avoids more specific adaptations as in [17, 24], which can lead to bases whose descriptions are more complex. Our adaptation of the basis also combines a heuristic for coefficient magnitude similar to that in [17], and a minimization of estimated $\text{RRMSE}(\hat{u})$ similar to that in [7], that is also novel to the authors' knowledge. We note that the BASE-PC method here should not be confused with the independently developed BASPC of [7], which has a similar acronym and purpose, as well as similarity in several computations. It also differs from the approach of [40] which focuses on identifying which dimensions are to be included into the approximation. A key difference between the approach here and other approaches is that the approach here is able to exploit sparsity, but does not explicitly depend upon it, and is capable of recovering both sparse and non-sparse solutions. It is suspected that many of the above methods too have this property, although this work demonstrates said property explicitly.

We also provide significant theoretical justification for the BASE-PC method, which can be possibly extended to other adaptive approaches. Under some justifiable assumptions we provide theoretical guarantees for both the basis and sample adaptive approaches used here. This analysis also expands to the case of non-sparse recovery which is a critical property for the basis adaptation approach, and fills a gap in analysis within the current basis adaptation literature. Further, we identify a set of functionals that under some assumptions are recovered by the BASE-PC method with a number of samples that does not depend on d , the dimension of the random inputs. In this case, the number of elements in the approximating basis also does not depend on d . This result is of interest with regards to the so-called curse-of-dimensionality associated with computations regarding high dimensional problems.

The organization of this paper has Section 2 describing the implementation of BASE-PC in detail with an algorithmic description of components critical for driving the basis and sample adaptations; Section 3 presenting numerical examples; and Section 4 providing theoretical justifications for the repeated iteration of the BASE-PC method.

2. BASE-PC Implementation Details

Here we present a detailed account of the BASE-PC iteration and its constituent functions presented in pseudocode, including default parameters. The implementation described here is that used in the examples of Section 3. These computations are divided into three categories corresponding to three subsections: Those computations associated with the evaluation and identification of the basis are presented in Section 2.1; those computations used for identification of new sample points are presented in Section 2.2; and those computations which identify the surrogate approximation for a given basis and sample set are presented briefly in Section 2.3. All of these components are utilized in a main iteration as described in Section 2.4.

2.1. Basis Evaluation and Update

For each input dimension, the identification of the one-dimensional orthonormal polynomials are given by the appropriate three-term recursion in a computationally efficient manner. We refer to this basic one-dimensional identification of a particular order by *basis_eval_1d*(**type**, p , ξ), where **type** determines the appropriate polynomial family; p refers to the maximal order polynomial to be computed in that dimension; and ξ refers to the point at which evaluation is occurring.

The identification of the multi-dimensional orthonormal polynomials is referred to as *basis_eval*(\mathcal{B}, ξ), where \mathcal{B} represents a description of the basis at which evaluation is occurring, including relevant order information, and ξ is the point at which the basis should be evaluated. This function identifies each one-dimensional evaluation via *basis_eval_1d*, before multiplying them appropriately to identify the evaluation of each basis function at the input.

It is necessary for bases of arbitrary anisotropic order to be constructed, and we refer to this function as *basis_id*(\mathbf{p}), where \mathbf{p} is as in (2), identifying the requested anisotropic basis. For brevity, a specific algorithm is not presented here, though the construction is explained in some detail relative to the construction of a total order basis.

First, the identification of the basis is done by sorting \mathbf{p} by dimension in a descending manner, so that $p_{(1)}$ corresponds to the maximal coordinate of \mathbf{p} . A loop is initialized so as to identify the total order basis of $p_{(1)}$, and each such basis function is tested to see if it meets the prescribed anisotropic order criteria. This determines whether or not the basis function is a member of the prescribed anisotropic total order basis, and it is added if it is a member.

Due to the sorting of orders, basis functions may be efficiently discarded, in that one failed test guarantees the failure of potentially many other basis functions, so that the number of tests is kept small. In this way, when $p_{(1)}$ is large but many other orders are small, relatively few basis functions need to be tested when compared to the potentially large size of the total order basis having potentially large order and dimension. Hence, the identification of the basis is computationally tractable even when the requested anisotropic order basis has a high order in some dimensions, and a large total number of dimensions of relatively low order. We note that in such a case iterating over the full total order basis associated with order $p_{(1)}$ would be infeasible due to the combinatorially large number of basis functions of a total order basis when both dimension and order are large. We note that this sorting of dimension based on the order of the anisotropic order basis is not kept for the remainder of what occurs, being used only for the construction of the basis.

For a given basis and set of input samples $\{\boldsymbol{\xi}^{(k)}\}_{k=1}^N$, we can form the measurement matrix $\boldsymbol{\Psi}$ that evaluates each basis function at each input, as in (6). With an additional weight matrix \boldsymbol{W} that is diagonal and positive-definite, we can form $\boldsymbol{D} = \boldsymbol{W}\boldsymbol{\Psi}$.

For a given basis, when the surrogate coefficients, \boldsymbol{c} , have been identified, we may remove m basis functions coinciding with small entries of \boldsymbol{c} . This allows us to shape and adapt the basis as per our heuristic of removing basis functions that have correspondingly small coefficient. We refer to this as basis contraction. We do this using a method called *basis_contract*($\mathcal{B}, \boldsymbol{c}, m$), and presented in Algorithm 1. The parameter m is looped over during the basis adaptation procedure. We note that in the case that multiple minimizing $|c_i|$ exist, we choose the one with smallest index i .

Algorithm 1: *basis_contract*($\mathcal{B}, \boldsymbol{c}, m$): Returns contraction of input basis, using information from a computed solution.

```

Set  $\mathcal{R} = \emptyset$  % Will contain basis functions to remove.
for  $k \leq m$  do
    Set  $k = \arg \min_{i \in \mathcal{B} \setminus \mathcal{R}} |c_i|$ . % Minimize over elements  $\mathcal{B}$  not in  $\mathcal{R}$ .
    Set  $\mathcal{R} = \mathcal{R} \cup \{k\}$ . % Add basis function to be removed.
end for
Return  $\mathcal{B} \setminus \mathcal{R}$  % Contracted basis is elements of  $\mathcal{B}$  not in  $\mathcal{R}$ .

```

Adjoint to contraction of the basis is expansion of the basis, through a function referred to as *basis_expand*(\mathcal{B}), and presented in Algorithm 2. We note that *basis_expand* expands general bases that do not coincide with anisotropic order bases, specifically bases that have had a number of basis elements removed via *basis_contract*. The parameter γ in *basis_expand* controls the relative expansion of the basis, with higher values leading to larger bases. For the examples in Section 3 $\gamma = 1.5$ is larger for the low dimensional problem of Sections 3.1 and $\gamma = 1.3$ is used for the low-dimensional problem in 3.4. Similarly, $\gamma = 1.01$ is smaller for the problems of Sections 3.2 and 3.3. The larger γ helps accelerate adaptation when the dimensions are smaller and the orders are expected to be relatively larger, while in the higher dimensional case it becomes more important to restrain the number of basis functions as the typical order of basis in any given dimension is low. Generally, small values of γ will work well, at the potential cost of needing more basis adaptation iterations.

We also include a certain number of new dimensions at order 1, denoted *dim_add*, which is set to 20 for the examples in Section 3. The modification for *dim_add* is most important for the example in Section 3.3. The other examples have 20 or fewer dimensions, and this constraint simply enforces that the minimal order in each dimension for those problems is 1, i.e. there is at least a linear term in each dimension.

<p>Algorithm 2: <i>basis_expand</i>(\mathcal{B}): Returns expansion of input basis.</p>

<p>Set $\mathbf{p} = \mathbf{0}$. % Will hold order information. for \mathbf{k} such that $\psi_{\mathbf{k}} \in \mathcal{B}$ do Set $\mathbf{p} = \max(\mathbf{p}, \mathbf{k})$. % Maximum is taken coordinate-wise. end for Add up to <i>dim_add</i> dimensions to \mathbf{p} at order 1. $\mathcal{B} = \text{basis_id}(\lceil \gamma \mathbf{p} \rceil)$. % Ceiling function is taken coordinate-wise.</p>

In the examples, *basis_contract* and *basis_expand* are used in tandem, repeatedly expanding further contracted bases. These contracted bases are further contracted by removing additional basis functions, leading to different expanded bases, and choosing the basis from a number of these by selecting which one produces a minimal validated error in surrogate approximation. As basis stability and obtaining the lowest available errors are a priority, it is reasonable to admit more solution solves. Hence, a basis can

be selected at each iteration from a set of candidate bases that minimizes an estimate of the RRMSE, a process which we refer to as basis validation. The algorithm to do this validation is summarized in Algorithm 3, and is referred to as *basis_validate*($\mathcal{B}_0, \mathbf{c}_0$). For the computations here, `max_strikes` = 6, where this parameter is instrumental for identifying the size of candidate bases we have to select from, where we stop identifying candidate bases with confidence that expansion of further contracted bases is unlikely to produce a basis with a lower estimate of RRMSE. Further, the basis adaptation procedure of expanding a contracted basis may be performed efficiently by noting that *basis_contract* need only remove one new element of an already sorted coefficient vector \mathbf{c} and new coefficients, and error estimates need only be computed when *basis_expand* produces a new basis. Here, a strike is an event where a validated error does not fall below the minimum achieved validated error. For computational efficiency the algorithm terminates if too many strikes are accumulated, resetting the strike counter if a new minimum is achieved.

Algorithm 3: *basis_validate*($\mathcal{B}_0, \mathbf{c}_0$): Returns validated basis from set of potential bases.

```

Let  $n = |\mathcal{B}_0|$ . % The number of basis elements in  $\mathcal{B}_0$ .
Set  $m = 0$ , strikes = 0, and min_error =  $\infty$ .
while  $m \leq n$  & strikes < max_strikes do
  Set  $\mathcal{B}_m = \text{basis\_expand}(\text{basis\_contract}(\mathcal{B}_0, \mathbf{c}, m))$ .
  if  $\mathcal{B}_m \neq \mathcal{B}_{m-1}$  then
    Evaluate all samples and QoI for  $\mathcal{B}_m$  to get  $\mathbf{D}_m$  and  $\mathbf{W}_m \mathbf{u}$ .
    Compute surrogate coefficients  $\mathbf{c}_m$  and estimate of RRMSE  $\epsilon_m$ .
    % Surrogate computation details are presented in Section 2.3.
    if  $\epsilon_m < \text{min\_error}$  then
      min_error =  $\epsilon_m$  & strikes = 0.
    else
      strikes = strikes + 1.
    end if
  end if
   $m = m + 1$ .
end while
Return basis achieving minimal validated error.

```

For cases of moderate dimensionality, the prescribed methods are sufficient. However when nearly linear scaling in dimension is required, it is useful to provide an upper bound on the orders prescribed for each dimension, a method referred to as *basis_upper_bound* and presented in Algorithm 4. This algorithm is only used for the example in Section 3.3, but is important there as without it, the number of basis functions during the basis expansion phase would quickly grow too large for tractable computation. We also note that this algorithm can be used by first ordering \mathbf{p} in descending order, although we do not do so here, as the dimensionality in Section 3.3 is already loosely sorted in a descending order of importance. This use of an upper bound

Algorithm 4: *basis_upper_bound*: Returns coordinate-wise upper bound on \mathbf{p} .

```

Let  $i_k$  index the last coordinate of  $\mathbf{p}$  having order  $k$ .
Initialize  $v$ 
Let  $k^*$  be max  $k$  such that  $i_k$  is defined.
for  $k \leq k^*$  do
    Set  $v_k = i_k + \text{dim\_add}$ .
end for
Initialize  $b$  % Is the vector that bounds the order in each coordinate.
for  $k \leq k^*$  do
    Set  $b(1 : v_k) = k$ . % Set first  $v_k$  entries of  $b$  to  $k$ .
end for
Set  $b(1 : \text{dim\_add}) = b(1 : \text{dim\_add}) + 1$ . % Increase order for first
dimensions.
```

on order at each iteration can prevent quadratic scaling in dimension from including 2nd order terms for a large number of dimensions. A linear or even constant approximation may be sufficient for most dimensions, and only a few dimensions need basis functions of higher order. Moreover these bounds may be systematically adjusted at each iteration, without a priori assumptions about an ideal basis for approximation. We note that another alternative to reduce the expansion of basis functions is to initialize m in Algorithm 3 to some integer greater than 0, although we do not consider doing so here. Adjusting this parameter would also reduce the size of expanded bases, and potentially reduce the number of bases for which estimates of the RRMSE need be computed.

After a solution has been updated in the new basis, we increase the number of samples used to compute coefficients. Motivated by a desire for a coherence-optimal sampling in our new basis, additional samples may be generated using the new basis as well as the basis used for sample generation in the previous iteration. This process is particularly useful in certain cases where high order approximations are needed in one or more dimensions, leveraging the benefits of coherence-optimal sampling [10, 11], and not requiring *a priori* knowledge about which dimensions require higher orders. Sometimes it is reasonable and practical to simply draw all samples from the same distribution, such as from the orthogonality distribution, and this provides a useful comparison for the examples in Section 3.

2.2. Sample Generation

In this work, when not sampling from an orthogonality distribution, sampling is done via Markov Chain Monte Carlo (MCMC) so as to minimize the coherence defined in [11]. We note that this distribution depends on the ℓ_2 -norm of the proposed vector of evaluated basis functions, as well as the orthogonality distribution. For each sample, denoted $\Xi^{(k)}$, a weight $w^{(k)}$ is associated, so that in aggregate the design matrix \mathbf{D} satisfies

$$\mathbb{E}(\mathbf{D}^T \mathbf{D}) = N\mathbf{I}. \quad (12)$$

For orthogonality distributions with infinite support, like the normal distribution, it is convenient to relax this requirement to holding only in an approximate sense [20, 10].

Our implementation for drawing N samples from a distribution g is referred to as *mcmc_sample(g,N)*. We note that this implementation of MCMC does not utilize adaptive proposal distributions, perpetually drawing proposals from the orthogonality distribution, though this is not ideal for e.g. high-order Hermite polynomials and the normal distribution [10]. Our method tunes the sampling with a burn-in parameter. Several burn-in samples are repeated until a running average of the normalization constant for the distribution is stabilized, as this helps to insure a quality sample, and then these burn-in samples are discarded and not utilized as draws from the desired distribution.

To improve the quality of sampling we also seek to limit the number of so-called collisions between samples, where a collision is defined to be when one MCMC sample is identical to the previous MCMC sample, which can

arise when a large number of potential samples are rejected in sequence. To prevent this we draw more intermediate samples before accepting the next sample. An upper bound on this collision rate is enforced, specifically $\exp(-8) \approx 0.00033$, and duplicate samples are not kept. This number, having no particular significance, may be reduced if more accurate samples are needed. This imparts a negligible bias in the MCMC sample as the collision rate may be kept quite low without much computational burden. We note that our parameters produce a quality sampling from most coherence-optimal distributions, while being computationally quick. However, a more careful sampling that utilizes more resources may produce better results. It is also possible to use these generated samples as candidates for more specific experimental designs [41, 42]. This design motivated approach is beyond the scope of this paper, and is a focus of future work.

The weight function $w(\boldsymbol{\xi})$ attached to every potential sample is related to the orthogonality distribution $f(\boldsymbol{\xi})$ and sampling distribution $g(\boldsymbol{\xi})$, as in [11]. For the initial sample,

$$g(\boldsymbol{\xi}) = c_g \|\boldsymbol{\psi}(\boldsymbol{\xi})\|_2^2 f(\boldsymbol{\xi}), \quad (13)$$

where $\boldsymbol{\psi}(\boldsymbol{\xi})$ is the row vector of realized basis functions evaluated at $\boldsymbol{\xi}$, $f(\boldsymbol{\xi})$ is the prescribed distribution for the uncertain inputs, and $c_g = |\mathcal{B}|^{-1}$ is the corresponding normalizing constant [11]. As

$$\mathbb{E}(\mathbf{D}^T \mathbf{D})_{i,j} = \int_{\Omega} w^2(\boldsymbol{\xi}) \psi_i(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) g(\boldsymbol{\xi}) d\boldsymbol{\xi},$$

it follows that (12) is satisfied when $w(\boldsymbol{\xi}) = \sqrt{|\mathcal{B}|} \|\boldsymbol{\psi}(\boldsymbol{\xi})\|_2^{-1}$.

2.2.1. Correction Sampling

At each BASE-PC iteration, we consider two bases. The previous basis, denoted \mathcal{B}_k , and the current basis, denoted \mathcal{B}_{k+1} . Each basis has an associated coherence-optimal distribution from Section 2.2, which we denote g_k and g_{k+1} , respectively. Our correction sampling assumes all previous samples were drawn from g_k , and wishes to draw additional samples maintaining (12), while having the aggregation of all samples be drawn in a way that resembles independent draws from g_{k+1} . This is done by implicitly defining the correction distribution g_k^c by the identity

$$(1 - \alpha_k) g_k(\boldsymbol{\xi}) + \alpha_k g_k^c(\boldsymbol{\xi}) = g_{k+1}(\boldsymbol{\xi}). \quad (14)$$

Here α_k must be chosen large enough such that $g_k^c(\xi) \geq 0$ for all ξ in the relevant domain. Additionally, considering g_{k+1} as a mixture of g_k and g_k^c , α_k is connected to the sample sizes from the previous basis, denoted N_k ; the new complete number of samples treated as if drawn from g_{k+1} , denoted N_{k+1} ; and the number of correction samples used to do this, denoted,

$$N_k^c := N_{k+1} - N_k.$$

Interpreting (14) in terms of this sampling idea,

$$\alpha_k = \frac{N_k^c}{N_{k+1}}.$$

These requirements are combined as outlined in Algorithm 5, which generates N_k^c transition samples, where N_k^c is identified within an acceptable range of values. This algorithm requires a few parameters. There is a parameter for maximum sampling ratio, denoted `max_sample_ratio` that enforces a maximum on how many correction samples are allowed as a ratio of the current sample size. In our examples, `max_sample_ratio` = 1, that is the sample size may at most double at each sampling. The primary benefit of setting `max_sample_ratio` is to not require an impractical number of correction samples. Also, there is a minimum sampling ratio `min_sample_ratio`, that bounds the minimum number of samples in the correction sample, relative to the current sample size, which is set a priori and varies for our examples between 0.1 to 0.3 depending on the computational budget. The main benefit of setting `min_sample_ratio` is to reduce the number of iterations that would occur if a low number of samples were generated on each iteration.

Algorithm 5 also specifies `weight_correction`, a variable that is used in the case that α in (14) must be chosen larger than what `max_sample_ratio` admits. Here, `weight_correction` artificially inflates α_k from (14) by giving samples from the correction distribution higher weight, producing an effect similar to having more samples from that distribution. The factor, `weight_correction`, is multiplied to all rows of \mathbf{D} corresponding to new samples generated by `sample_expand`, i.e. associated with the correction sampling. Its primary role is to insure that (12) holds after the correction sampling. This multiplication is done for the next solution computation only, and for all subsequent samples the generated random variables are all assumed to have been drawn independently from the prescribed g_k .

There are two reasons for this. First, the correction sampling assumes all previous samples are drawn from g_k , and maintaining previous weights

contradicts this assumption. Further, for any given iteration, the violation of (12) that comes from misrepresenting previous weights vanishes as the overall sample size increases. Second, it is preferable that the aggregate sample not maintain lasting effects from previous samples. Having a few previous samples from a correction sampling that had attached to it a very large weight would potentially lead to the function being fit unnecessarily well at those points, at the detriment of other points in the domain. Stated another way, the RRMSE in the surrogate would be increased by inappropriately fitting some areas of the domain due to the persistence of weights.

Algorithm 5: *sample_expand*($\mathcal{B}_k, \mathcal{B}_{k+1}$): Returns sample with correction to be used for next solution computation.

```

Set  $\alpha_k = \text{min\_sample\_ratio}$ .
while  $\alpha_k$ -validated sample not generated do
  Set  $N_k^c = \lceil \alpha N_k \rceil$ .
  Set  $\alpha_k = N_k^c / (N_k + N_k^c)$ . % Ceiling function changes  $\alpha$  slightly.
  Define  $g_k^c$  via (14).
  Set  $(\tilde{\alpha}_k, \text{sample}) = \text{mcmc\_sample}(g_k^c, N_k^c)$ . %  $\alpha_k$  may be increased.
  % The need to increase  $\alpha_k$  is revealed during sampling.
  if  $\tilde{\alpha}_k > \alpha_k$  then
    Set  $\alpha_k = \tilde{\alpha}_k$ . % Increase  $\alpha_k$  if needed.
  else
    Break while loop % Here  $\alpha_k$  has validated on the sample.
  end if
end while
if  $\alpha_k > \text{max\_sample\_ratio}$  then
  Remove samples from  $N_k^c$  so that  $N_k^c / N_k < \text{max\_sample\_ratio}$ .
  Set  $\alpha'_k = N_k^c / (N_k + N_k^c)$ . % Note that  $\alpha'_k < \text{max\_sample\_ratio}$ .
  Set  $\text{weight\_correction} = \alpha_k^{-1} \alpha'_k$ . % This is larger than 1.
  Set  $\text{true\_sample\_ratio} = \alpha'_k$ .
else
  Set  $\text{weight\_correction} = 1$ . % No weight correction necessary.
  Set  $\text{true\_sample\_ratio} = \alpha_k$ .
end if

```

2.3. Surrogate/Coefficient Identification

With a basis and sample identified, a surrogate solution is identified by computing coefficients for each basis function. These coefficients are computed by solving (7) with a cross-validated δ [9] to minimize a validated estimate of RRMSE, using a certain number of folds and a certain number of validation samples in each fold. Here the range of δ is given based on the previous validated error or an initial value. Specifically, the set of potential δ is 0 and a set of 20 tolerances that are spaced, evenly in a logarithmic scale, around the largest of the previous minimizing tolerance or validated error. Further, 24 randomly generated partitions of the data are used to compute an estimate of the RRMSE and a corresponding δ for each partition. For each such partition, 80% of samples are used for computation of the solution, while 20% are used for validation. The number of partitions and percentage of validation samples are generally larger than needed for a relatively accurate estimation of error. We note that the method of error estimation used here is closely related to the leave-one-out error estimate of [7]. It may be useful in certain situations to consider other validation techniques, although this is sufficient for the examples here.

2.4. Main Iteration

The main iterative process then is to sequentially identify new bases for the surrogate approximation and new samples that are compatible with this sequence of bases so that the aggregate sample at each iteration mimics a coherence-optimal sample for the appropriate basis at that iteration. To clarify the presentation, we define *initialize* as a function that produces some initial basis; a number of samples that are coherence-optimal for that basis; surrogate coefficients for that basis; and an estimate of RRMSE. For our examples, we initialize to a total-order basis with some small number of samples drawn from the coherence-optimal distribution, unless all samples are being drawn from the orthogonality distribution. The surrogate coefficients and RRMSE estimate are then computed in that basis for those samples as by the method described in Section 2.3. The BASE-PC algorithm is then described in Algorithm 6, and referred to as *base-pc-loop*. Here `max_iterations` may be set based on convergence criteria. For our examples, the loop is run until computational time grows large, although it is also reasonable to stop based on the RRMSE estimates as generated by *basis-validate*.

<p>Algorithm 6: <i>base-pc-loop</i>: The main iteration for BASE-PC.</p> <p>Set $(\mathcal{B}_0, \mathbf{c}_0, \mathcal{S}) = \text{initialize}()$. % We let \mathcal{S} denote identified samples.</p> <p>for $k = 1 : \text{max_iterations}$ do</p> <p> $(\mathcal{B}_k, \mathbf{c}_k) = \text{basis_validate}(\mathcal{B}_{k-1}, \mathbf{c}_{k-1})$.</p> <p> $\mathcal{S} = \text{sample_expand}(\mathcal{B}_k, \mathcal{B}_{k-1})$.</p> <p>end for</p>

3. Numerical Examples

To investigate the numerical efficacy of the BASE-PC iteration, we investigate four problems. The first in Section 3.1 is a low-dimensional smooth problem that is traditionally targeted for interpolation and regression problems. The second in Section 3.2 is a moderate dimensional problem with some characteristic coefficient decay often seen in engineering problems. The third in Section 3.3 is a 1000 dimensional manufactured problem that shows the BASE-PC method can be effective at dimensions not usually associated with PC accuracy. The final example in Section 3.4 is a low dimensional surface adsorption model that is not well suited to polynomial approximation, having many properties that may preclude it from use with PC, but demonstrating BASE-PC's improvement when polynomial approximation is of suspect accuracy, as occurs in many practical problems.

In all examples here the total order bases use only samples drawn from the orthogonality distribution as opposed to any coherence-optimal sampling. For the BASE-PC methods, sample adaptivity (SA) refers to use of the correction sampling distribution with coherence-optimal sampling as in [11], while no sample adaptation (No SA) refers to using samples from the orthogonality distribution. In both cases, basis adaptation is performed in the same manner. In all cases validated RRMSE represents the estimated RRMSE as identified by BASE-PC, while RRMSE is a reference estimate of RRMSE computed using a large number of independently generated samples.

3.1. Case I: Franke function

One function that is often used in regression or interpolation analysis is the Franke function [43], which is a two dimensional function defined on

$[0, 1] \times [0, 1]$ by

$$\begin{aligned}
u(\Xi) := & \frac{3}{4} \exp \left(-\frac{(9\Xi_1 - 2)^2}{4} - \frac{(9\Xi_2 - 2)^2}{4} \right) + \frac{3}{4} \exp \left(-\frac{(9\Xi_1 + 1)^2}{49} - \frac{9\Xi_2 + 1}{10} \right) \\
& + \frac{1}{2} \exp \left(-\frac{(9\Xi_1 - 7)^2}{4} - \frac{(9\Xi_2 - 3)^2}{4} \right) - \frac{1}{5} \exp \left(-(9\Xi_1 - 4)^2 - (9\Xi_2 - 7)^2 \right),
\end{aligned}
\tag{15}$$

and depicted in Figure 1. The results of running the BASE-PC iterations are shown in Figure 2, demonstrating improvement for adapted bases over the use of total order bases in that each of the total order bases are only as accurate as the adaptive bases for a range of sample sizes. Specifically, when comparing the number of QoI evaluations to the RRMSE, we see that the sample adaptive BASE-PC iterations reliably outperform other methods, and that the BASE-PC iterations with and without sample adaptation use significantly fewer basis functions for the same level of accuracy when compared to the total order bases, and require no *a priori* information about what order of basis to utilize.

These plots show a gradual increase of the average number of basis functions included, as the number of QoI evaluations increases with the BASE-PC approach, a common theme in all the examples. This example also shows the benefit of sample adaptation, as higher sample sizes allow more exceptional accuracy when sample adaptation is performed. We note that in this case the order of adapted approximation remains comparable in both dimensions, so that the basis adaptivity behaves similarly to identifying a particular total order approximation.

3.2. Case II: Stochastic heat driven cavity flow

A practical case for consideration comes from temperature driven fluid flow in a cavity [44, 2, 45, 29], where the QoI is a component of the velocity field at a fixed point and time. The left vertical wall has a uniform temperature \tilde{T}_h , referred to as the hot surface, while the right vertical wall has a variable temperature \tilde{T}_c , and is referred to as the cold surface; both walls are adiabatic. The reference temperature is defined as $\Delta\tilde{T}_{ref} := \tilde{T}_h - \tilde{T}_c$. Let $\hat{\mathbf{y}}$ denote the unit normal vector in the vertical dimension. The non-dimensional

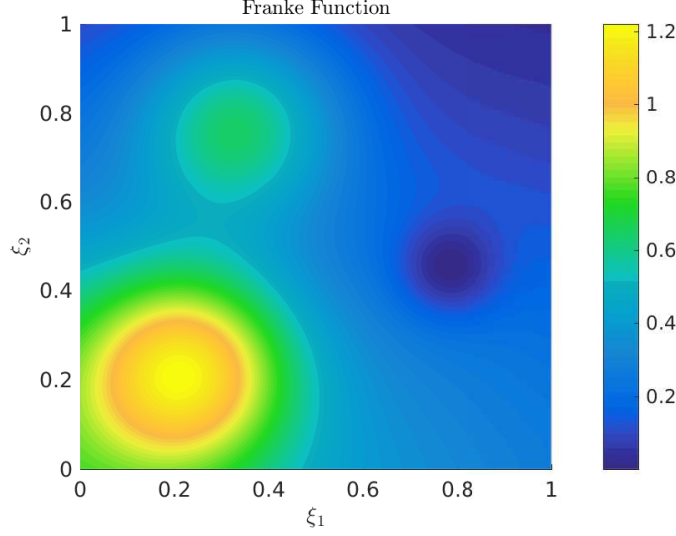


Figure 1: The Franke function.

governing equations are given by

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} &= -\nabla p + \frac{\text{Pr}}{\sqrt{\text{Ra}}} \nabla^2 \mathbf{u} + \text{Pr} T \hat{\mathbf{y}}, \\
\nabla \cdot \mathbf{u} &= 0, \\
\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{u} T) &= \frac{1}{\sqrt{\text{Ra}}} \nabla^2 T,
\end{aligned} \tag{16}$$

where \mathbf{u} is velocity vector, p is pressure, T is normalized temperature and t is time. Non-dimensional Prandtl and Rayleigh numbers are defined, respectively, as $\text{Pr} := \tilde{\mu} c_p / \tilde{\kappa}$ and $\text{Ra} := \tilde{\rho} \tilde{g} \beta \Delta \tilde{T}_{ref} \tilde{L}^3 / (\tilde{\mu} \tilde{\kappa})$ where tilde denotes dimensional quantities: $\tilde{\rho}$ is density, \tilde{L} is reference length, \tilde{g} is gravitational acceleration, $\tilde{\mu}$, is molecular viscosity and $\tilde{\kappa}$ are is thermal conductivity. The coefficient of thermal expansion is $\beta = 0.5$. In this example the Prandtl and Rayleigh numbers are given by $\text{Pr} = 0.71$ and $\text{Ra} = 10^6$.

3.2.1. Stochastic Boundary Conditions

On the cold wall, a temperature distribution with stochastic fluctuations is applied,

$$T(x = 1, y) = T_c + T'(y), \tag{17}$$

where $T_c = -0.5$ is a constant expected temperature, and $T_h = 0.5$ is the temperature on the hot wall. The fluctuation $T'(y)$ is given by the truncated Karhunen-Loève expansion

$$T'(y) = \sigma_T \sum_{i=1}^d \sqrt{\lambda_i} \varphi_i(y) \Xi_i, \quad (18)$$

where $d = 20$ and $\sigma_T = 11/100$. Here, each Ξ_i is assumed to be an independent and identically distributed uniform random variable on $[-1, 1]$, with $\{\lambda_i\}_{i=1}^d$ and $\{\varphi_i(y)\}_{i=1}^d$ the d largest eigenvalues and the corresponding eigenfunctions of the exponential covariance kernel

$$C_{TT}(y_1, y_2) = \exp\left(-\frac{|y_1 - y_2|}{l_c}\right), \quad (19)$$

where $l_c = 1/21$ is the correlation length. An example of cold boundary condition is shown in figure 3b. Our QoI is the vertical velocity component at $(0.25, 0.25)$. The QoI computations here do not solve this model directly, but instead use a surrogate solution computed using a basis of 2500 elements reduced from a total order 4 basis and a large number of samples.

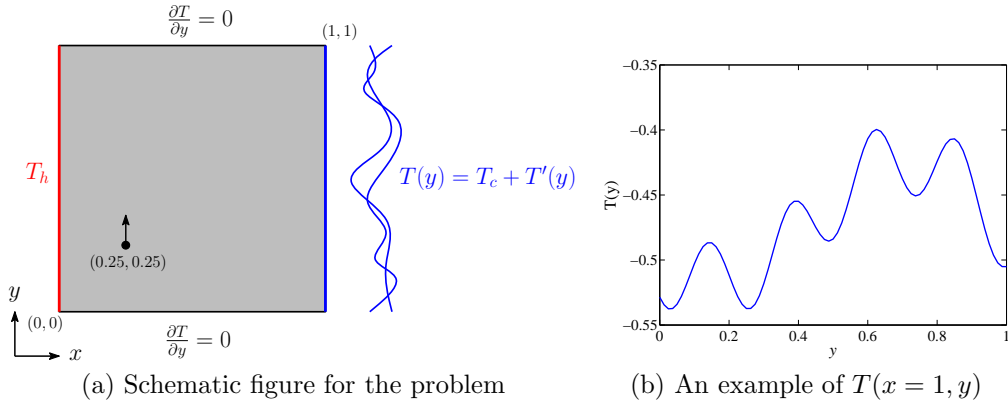


Figure 3: Illustration of the problem, reproduced from Figure 6 of [29].

3.2.2. BASE-PC Iterations

The results of running the BASE-PC iterations are shown in Figure 4, demonstrating dramatic improvement for adapted bases over the use of total order bases. This improvement is seen even when no sample adaptivity is

done, i.e. when all samples are drawn from the orthogonality distribution. We note that the number of adapted basis elements is correlated strongly to the number of QoI evaluations, and that the correlation between the validated RRMSE and the actual RRMSE is also high for all methods. This problem is smooth in the input parameters, which facilitates an easy basis adaptation and leads to a smooth decay in RRMSE as the number of QoI evaluations increases for the basis adaptive methods. The non-adaptive total order bases are not tuned to the number of samples, nor distribute basis functions ideally between dimensions leading to a recovery with reduced effectiveness.

3.3. Case III: 1000-Dimensional Manufactured Decay

As a demonstration of scaling for a high dimensional problem, consider

$$u(\Xi) = \exp \left(2 - \sum_{k=1}^d \frac{\sin(k)\Xi_k}{k} \right), \quad (20)$$

with $d = 1000$. Here each Ξ_k is independent and uniformly distributed on $[0, 1]$. For computations at this dimensionality, order control is implemented for the basis expansion so that instead of increasing each p_i , a limited number of dimensions have increased p_i at each iteration as dictated by Algorithm 4. Specifically, we set `dim_add` = 20.

The results in Figure 5 are a computation for a non-linear polynomial approximation in 1000 dimensions. We notice that the adaptive methods still exhibit a smooth reduction in RRMSE with regards to the number of QoI evaluations, although the rate of this reduction is not as large as that for the cavity flow problem in Section 3.2.2. This is coupled with a high correlation between the number of basis functions and QoI evaluations, as well as the estimated RRMSE and an accurate reference RRMSE.

We may also consider how this method compares to Monte Carlo estimation of the first two moments of the distribution, given that this is a widely used approach for problems of this dimensionality. In Figure 6 we see the comparison of errors in the mean and variance computations for this problem. We note that the BASE-PC iterations are generally more accurate in estimating the mean and variance than corresponding Monte Carlo computations, although the regression (7) reduces $\|\mathbf{c}\|_1$ and does produce a bias to underestimate these quantities. This bias is negligible at larger sample sizes, but is significant at smaller sample sizes. Overall the BASE-PC moment estimates have lower error, and also have the benefit of producing a surrogate model that explains much of the variance in addition to estimating it.

3.4. Case IV: Surface Adsorption

While the previous examples are generally smooth and well approximated by low order polynomials, some QoI have stiff response to the input randomness and require high degree polynomials. One such problem is to quantify the uncertainty in the solution ρ of the non-linear evolution equation

$$\begin{cases} \frac{d\rho}{dt} = \alpha(1 - \rho) - \gamma\rho - \kappa(1 - \rho)^2\rho, \\ \rho(t = 0) = 0.9, \end{cases} \quad (21)$$

which models the surface coverage of certain chemical species, as examined in [46, 47]. We consider uncertainty in the adsorption, α , and desorption, γ , coefficients, and model them as shifted log-normal variables. Specifically, we assume

$$\begin{aligned} \alpha &= 0.1 + \exp(10 \Xi_1), \\ \gamma &= 0.001 + 0.001 \exp(10 \Xi_2), \end{aligned}$$

where we consider Ξ_1, Ξ_2 as standard normal $\mathcal{N}(0, 1)$ random variables; hence, the dimension of our random input is $d = 2$. We note that this example differs from the corresponding example in [11], which has $0.05 \Xi_1$ and $0.05 \Xi_2$ in place of $10 \Xi_1$ and $10 \Xi_2$ in the arguments of the exponentials. Also, the 0.001 parameter multiplying γ differs from 0.01 in [11] which in this work somewhat reduces the relative variability with respect to γ when compared to that of α . In aggregate, the example here corresponds to significantly higher uncertainty in the input parameters. The reaction rate constant κ in (21) is assumed to be deterministic and is set to $\kappa = 10$.

Our QoI is $\rho_c := \rho(t = 4, \Xi_1, \Xi_2)$, and to approximate this, we consider a Hermite PC expansion in the two variables, Ξ_1 and Ξ_2 . This problem is interesting for its stiff transition and the need for high order polynomials in Ξ_1 , with a lower but still considerable order of polynomials in Ξ_2 .

For normally distributed input random variables, we utilize the associated Hermite polynomials for the approximation, which are known to be more difficult to use with polynomial approximation [10, 11]; specifically, regression via (7) using Hermite polynomials is quite sensitive for high order approximations. When high order Hermite polynomials are used, sampling from the orthogonality distribution is ineffective due to a dependence on exceptionally rare events for accurate approximations. The lack of smoothness in this problem is generally exacerbated by using Hermite polynomials, with

results shown in Figure 7. We note that while no method does particularly well for this problem, the BASE-PC method, specifically the sample adaptive version, significantly outperforms the total order computations.

Figure 8 shows the realizations of the true model QoI and the realizations from a BASE-PC surrogate constructed using sample adaptation, where the realizations of the BASE-PC surrogate are set to 0 if they are negative values, and set to 1 if they are greater than 1. This constraint is done as the physical model dictates values be in $[0, 1]$, and this change makes the plots in Figure 8 comparable as the BASE-PC surrogate takes values up to approximately 8 and down to approximately -1 . The figure shows that this QoI exhibits behavior that makes it very difficult to approximate by a polynomial, and indeed the utility of doing so for practical purposes is suspect. Here, we consider it as a contrast to the Franke function in Section 3.1. While the Franke function is smooth and well approximated by polynomials of modest order, this function has sharp transitions that approach discontinuity, and large areas of effectively no variation, which is a function that is not well approximated by polynomials; even here where polynomial orders in Ξ_1 reach the hundreds. The surrogate is noticeably deficient around the edge of the transition, and exhibits inaccuracy in approximating the constant regions, both of which are consistent with polynomial approximations to functions of this type.

We also note that the use of the orthogonality distribution as a proposal distribution is lacking for an accurate coherence-optimal sampling here, tending to generate samples further towards the origin than an accurate coherence-optimal sampling. A more accurate sampling may improve solution recovery when utilizing the sample adaptive approach by better leveraging these rare events. Of some interest is noting that the reference RRMSE and validated RRMSE are significantly less correlated in this example, due to both estimates being significantly less accurate for this problem, though the BASE-PC method, particularly the sample adaptive version, is more accurate than the total order versions. We note that the use of sampling from the orthogonality distribution for proposal samples in the MCMC leads to the coherence-optimal sampling here being significantly less accurate than that of [11].

4. Theoretical Exposition

Here we present theoretical justification for the BASE-PC iteration, particularly with regards to the iterative basis adjustment and correction sampling. We also address the recovery via ℓ_1 -minimization and its analysis that depends on sparsity, particularly as the goal of adapting a basis is to have contributions from as many basis functions as possible, i.e. to identify representations that are not sparse. However, it is still of great practical utility to be able to recover sparse solutions in an expanded basis.

We present this analysis in the following sections. Section 4.1 details some specifics of the coherence based approach we rely on here, as well as a notion of coupling that is key to the analysis of basis adaptation and some useful matrix bounds for the remaining sections. Section 4.2 handles the recovery results for sparse solutions. Section 4.3 details recovery results that are useful for recovering solutions that we consider non-sparse. Section 4.4 identifies a class of problems that under some assumptions may be recovered in a number of samples and basis functions that is independent of dimension.

4.1. Preliminaries

Here we present some of the preliminaries used for our main results, including concepts, notation and a few results used in the remaining sections. We first note that all results here rely on a noise model that is at least with high probability uniformly bounded. Let $\psi(\xi)$ denote the realized row vector that evaluates the basis functions in \mathcal{B}_k at ξ ; let $w(\xi)$ denote the weight associated with the coherence-optimal sampling associated with \mathcal{B}_k , and let $\epsilon_{\mathcal{B}_k}$ denote the truncation error associated with the basis \mathcal{B}_k as from (11). For us, we then require that $\|\epsilon_{\mathcal{B}_k}(\xi)w(\xi)\psi(\xi)\|_\infty \leq \lambda$ holds with high probability for some λ . We note that this is not a problem for most $u(\xi)$, and when using most practical distributions for ξ and the corresponding orthogonal polynomials and coherence-optimal weights, but could be an issue in more exotic cases.

4.1.1. Sparse vs. Non-Sparse Recovery

The BASE-PC method relies on the basis adaptation procedure to maintain a ratio of samples to basis functions that admits recovery, and we seek to identify a ratio of samples to basis functions that guarantees this stability. The basis adaptivity muddles the compressed sensing interpretation in that we actively seek to have the sparsity parameter be large relative to the number of basis functions. Because of this, we present our main results in both

sparse and non-sparse cases. Here and throughout s represents the sparsity parameter, and is broadly the number of non-zero coefficients needed to recover a QoI with a particular polynomial basis. Generally, s increases as more accurate solutions are requested, and the relationship between s and the accuracy of a solution is not addressed here. We do note that the tradeoff with s can be partially interpreted in terms of truncation error as in (11), and this does show up in the presented results.

Heuristically, if basis adaptation is successful then we generally expect a large fraction of basis functions to be useful for recovering the QoI, and this falls into the non-sparse recovery framework. It is also useful to insure that when there is a relatively small number of useful basis functions in our basis, that we may still have a quality recovery, and this recovery is referred to as sparse recovery. Another benefit of sparse recovery is with regards to how aggressively one may expand a basis at each iteration. An ability to accurately recover sparse reconstructions implies that we may add in a relatively large number of basis functions where few of them are expected to be useful in approximating the QoI. In terms of the BASE-PC method discussed here, this means that larger γ and `dim_add` may be used in the *basis_expand* algorithm of Algorithm 2. If the number of necessary basis functions is $s < 0.5|\mathcal{B}_k|$, i.e. s is less than half the number of basis functions used at the k th iteration, then we consider this to be sparse recovery, and otherwise we consider s to be non-sparse recovery. The surrogate identification may also be anticipated to be more robust with regards to the expansion and contraction parameters that determine the basis adaptation.

We note that sparse recovery has additional factors in $\log(s)$, that are unnecessary for the non-sparse recovery result, and if the non-sparse results are more favorable then those may be utilized freely, as may occur for s near $0.5|\mathcal{B}_k|$. We also note that there are some nuances that exist in the case of very sparse solutions; but that there is no issue with any of the results if we assume $s \geq 2$, and results can be defined for $s \geq 1$ with some changes to the presentation. We briefly remark on this later in Section 4.1.4. There is also some beneficial improvements to the number of samples when the sparsity is very high, such as when the adapted basis is of high quality and s approaches $|\mathcal{B}_k|$, on which we also remark later in Section 4.1.4. The reason for both of these results arise from the probabilistic approach to relevant bounds on the design matrix which we discuss over the course of the next several sections.

4.1.2. Coherence

Let Ω be the domain of the random input being considered, \mathcal{B} the current basis, and $\boldsymbol{\psi}(\boldsymbol{\xi})$ a $1 \times |\mathcal{B}|$ vector whose entries are the evaluation of the basis functions ψ_k at $\boldsymbol{\xi}$. Let $w(\cdot)$ denote the weight function associated with the importance sampling that determines how the $\boldsymbol{\xi}$ are drawn, so that $w(\boldsymbol{\xi})$ is the weight function evaluated at $\boldsymbol{\xi}$. Consider the definitions,

$$\mu_\infty := \max_{\boldsymbol{\xi} \in \Omega} \|w(\boldsymbol{\xi})\boldsymbol{\psi}(\boldsymbol{\xi})\|_\infty^2; \quad (22)$$

$$\mu_2 := \max_{\boldsymbol{\xi} \in \Omega} \|w(\boldsymbol{\xi})\boldsymbol{\psi}(\boldsymbol{\xi})\|_2^2; \quad (23)$$

$$\mu_2(s) := \max_{\boldsymbol{\xi} \in \Omega} \max_{|\mathcal{S}| \leq s} \sum_{k \in \mathcal{S}} |w(\boldsymbol{\xi})\psi_k(\boldsymbol{\xi})|^2. \quad (24)$$

These represent the potential maximum of certain vector norms over potential rows in the matrix \mathbf{D} , where the coherence-optimal importance sampling is to minimize this maximum. We note that as in [20, 10], the set Ω could be truncated if for example, these maximums are not bounded over the domain Ω , as occurs with e.g. Hermite polynomials. While μ_∞ has been used within the context of ℓ_1 -minimization [20, 10, 48], and μ_2 has been used within the context of ℓ_2 -minimization [19, 11, 49], it may be more appropriate to consider $\mu_2(s)$ in the case of ℓ_1 -minimization. We note that the importance sampling of [11] which is used here insures that $\mu_2 = |\mathcal{B}|$, the minimal possible value attainable by independent sampling. We note too that $\mu_2 = \mu_2(|\mathcal{B}|)$. Further, straightforward bounds can be found relating these notions, as summarized by the following lemma.

Lemma 1. *With the coherence parameters defined as in (23) it follows that,*

$$\begin{aligned} \max \left(\frac{s}{|\mathcal{B}|} \mu_2, \mu_\infty \right) &\leq \mu_2(s) \leq \min(\mu_2, s\mu_\infty); \\ \max \left(\frac{s}{|\mathcal{B}|} \mu_2, \mu_2(s) \right) &\leq s\mu_\infty \leq \min(s\mu_2, s\mu_2(s)); \\ \max(\mu_\infty, \mu_2(s)) &\leq \mu_2 \leq \min \left(|\mathcal{B}| \mu_\infty, \frac{|\mathcal{B}|}{s} \mu_2(s) \right). \end{aligned}$$

Proof. These results follow from standard inequalities of vector norms. ■

The quantities in the center of the inequality chain may each be used to bound ℓ_1 -recovery of a solution of sparsity $\lfloor s/2 \rfloor$ with similar bounds, in a

manner compatible with the analysis of [20]. We denote any definition in the center of the above inequalities; one of $\mu_2(s)$, $s\mu_\infty$, or μ_2 ; by μ , and for simplicity of presentation we focus on $\mu = \mu_2(s)$ in what follows. The ℓ_2 -coherence-optimal sampling used in the examples here, is optimal with regards to minimizing μ_2 over all independent random sampling distributions [11]. Note that $\mu_2(s)$ is the smallest of these three, but that $\mu_2(s)$ involves a maximum over a combinatorially large set $\{|\mathcal{S}| \leq s\}$, that complicates the analysis. Interestingly, it is simple enough to perform a coherence-optimal sampling that minimizes $\mu_2(s)$, as for any realized candidate vector, $\psi(\xi)$, the weight function, $w(\xi)$, and hence the MCMC sampling, involves only identifying the s elements of the candidate row that have the largest absolute value. However, such a sampling is beyond the scope of this work, but could be useful in cases where a sparsity parameter s is either assumed *a priori* or estimated in some manner.

4.1.3. Matrix Bounds

Here we present matrix bounds that will be used to show our results in Section 4.2 and Section 4.3. Before presenting a key matrix bound that will be used to justify our coherence-optimal sampling to minimize μ_2 , as well as our associated correction sampling, we discuss a normalization for the design matrix, denoted \mathbf{D} that is made throughout. We utilize a bound in a probabilistic sense of the quantity $\|\mathbf{D}^T \mathbf{D} - \mathbb{E}(\mathbf{D}^T \mathbf{D})\|$, where in what remains, all unspecified matrix and vector norms are assumed to be ℓ_2 -norms. To consider the convergence of $\mathbf{D}^T \mathbf{D}$ to its mean in terms of the sample size, N , we normalize $\mathbf{D}^T \mathbf{D}$ so that,

$$\mathbb{E}(\mathbf{D}^T \mathbf{D}) = \mathbf{I}, \quad (25)$$

or at a minimum we require that this holds approximately. We note that division of \mathbf{D} by a constant is associated with a similar normalization on $\mathbf{W}\mathbf{u}$, and that there is no effect on the computed surrogate when this is accounted for. That is, this normalization is a theoretical convenience with no effect on the computed solution or its associated error. We also note that this normalization differs from that in (12), which would be inconvenient here.

For our purposes we let \mathcal{S} denote a subset of the basis having size $|\mathcal{S}|$. We use the subscript of \mathcal{S} to denote that the associated matrix is restricted to only those entries relevant for basis functions in \mathcal{S} .

The next probabilistic matrix bound is of a type that is useful for guaranteeing recovery of accurate, stable function approximations [50, 51, 48, 20, 10, 11]. Specifically, we cite results of Section 5.4 of [51], with Theorem 5.44 of that work being directly applied here. We present that theorem here in a slightly different form, and as a lemma.

Lemma 2. [51] *Let*

$$\mathbf{E}_S := \mathbb{E}(\mathbf{D}_S^T \mathbf{D}_S).$$

There exists $\kappa > 0$ depending only on $\|\mathbf{E}_S\|^{-1/2}$, such that

$$\mathbb{P}\left(\|\mathbf{D}_S^T \mathbf{D}_S - \mathbf{E}_S\| > t\right) \leq |\mathcal{S}| \exp(-\kappa t N \mu^{-1}).$$

Proof. This is a rearrangement of Theorem 5.44 of [51] noting that in the context of that theorem, where \mathbf{A}_i corresponds to the i th row of \mathbf{D}_S , $\|\mathbf{A}_i\|_2 \leq \sqrt{\mu}$ almost surely for all i . ■

We also show that samples generated from the correction sampling of Section 2.2.1 will complement the old samples in a way such that (25) holds, and the realized $\mathbf{D}_S^T \mathbf{D}_S$ is near its mean. We first prove this for an individual iteration of sampling. Recall that the purpose of the correction sampling is not to alter moments, but to maintain low aggregate coherence in the samples. For what remains we assume that `max_sample_ratio` from Section 2.2.1 is set to infinity. This simplifies much of what follows, specifically we avoid technical complications that would arise from considering `weight_correction`.

Lemma 3. *Let $\mathbf{D}_{S,1}$ be the design matrix associated with an initial set of N_1 samples, and $\mathbf{D}_{S,2}$ that with a correction sampling as from Section 2.2.1 using N_2 samples. Let $\mathbf{E}_{S,1}$ and $\mathbf{E}_{S,2}$ denote the expectations of $\mathbf{D}_{S,1}^T \mathbf{D}_{S,1}$ and $\mathbf{D}_{S,2}^T \mathbf{D}_{S,2}$, respectively. Let \mathbf{D}_S be the full design matrix, restricted to those entries relevant to \mathcal{S} that are used for the computation of \hat{u} . For any fixed t ,*

$$\mathbb{P}\left(\|\mathbf{D}_S^T \mathbf{D}_S - \mathbf{I}\| > t\right) \leq |\mathcal{S}| \min_{\tau_1 + \tau_2 = t} \left(\exp(-\kappa_1 \tau_1 N_1 \mu_1^{-1}) + \exp(-\kappa_2 \tau_2 N_2 \mu_2^{-1}) \right),$$

where κ_i depends only on $\|\mathbf{E}_{S,i}\|^{-1/2}$ and μ_i is associated with samples from the corresponding distribution.

Proof. Applying Lemma 2 to $\mathbf{D}_{\mathcal{S},1}$ and $\mathbf{D}_{\mathcal{S},2}$, implies that

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{D}_{\mathcal{S},1}^T \mathbf{D}_{\mathcal{S},1} - \mathbf{E}_{\mathcal{S},1}\| > \tau_1\right) &\leq |\mathcal{S}| \exp\left(-\kappa_1 \tau_1 N_1 \mu_1^{-1}\right); \\ \mathbb{P}\left(\|\mathbf{D}_{\mathcal{S},2}^T \mathbf{D}_{\mathcal{S},2} - \mathbf{E}_{\mathcal{S},2}\| > \tau_2\right) &\leq |\mathcal{S}| \exp\left(-\kappa_2 \tau_2 N_2 \mu_2^{-1}\right),\end{aligned}$$

where $\mathbf{E}_{\mathcal{S},1}$ and $\mathbf{E}_{\mathcal{S},2}$ are the associated expectations so that by the construction in Section 2.2.1, $\mathbf{E}_{\mathcal{S},1} + \mathbf{E}_{\mathcal{S},2} = \mathbf{I}$. Recall that μ_1 and μ_2 are the coherence parameters associated with the differing samples. Noting that

$$\begin{aligned}\mathbf{D}_{\mathcal{S}}^T \mathbf{D}_{\mathcal{S}} &= \mathbf{D}_{\mathcal{S},1}^T \mathbf{D}_{\mathcal{S},1} + \mathbf{D}_{\mathcal{S},2}^T \mathbf{D}_{\mathcal{S},2}; \\ \mathbf{D}_{\mathcal{S}}^T \mathbf{D}_{\mathcal{S}} - \mathbf{I} &= \mathbf{D}_{\mathcal{S},1}^T \mathbf{D}_{\mathcal{S},1} - \mathbf{E}_{\mathcal{S},1} + \mathbf{D}_{\mathcal{S},2}^T \mathbf{D}_{\mathcal{S},2} - \mathbf{E}_{\mathcal{S},2}; \\ \|\mathbf{D}_{\mathcal{S}}^T \mathbf{D}_{\mathcal{S}} - \mathbf{I}\| &\leq \|\mathbf{D}_{\mathcal{S},1}^T \mathbf{D}_{\mathcal{S},1} - \mathbf{E}_{\mathcal{S},1}\|_2 + \|\mathbf{D}_{\mathcal{S},2}^T \mathbf{D}_{\mathcal{S},2} - \mathbf{E}_{\mathcal{S},2}\|_2,\end{aligned}$$

completes the lemma. ■

Lemma 3 is somewhat unsatisfying in that there is no guarantee that μ_1 and μ_2 are well behaved, even though this is the key heuristic behind the correction sampling. Still, Lemma 3 communicates that a convergence of the gramian of the design matrix to identity is maintained by correction sampling.

We note here that the use of multiple correction samplings leads to a sum of exponentials in Lemma 3. Given the difficulty in addressing the individual sampling coherences, showing a convergence with this method as the number of iterations increases would be difficult. Instead of this, we argue differently, using the technology of coupling presented in Section 4.1.4. Before that however, we present one final result that we utilize when showing uniform recovery in the sparse case, as well as use for the non-sparse case. This also introduces key notation that is used in our approximation results, as well as demonstrating a key result for the recovery of solutions via ℓ_1 -minimization.

Let \mathcal{B} denote the basis at a fixed iteration of BASE-PC. Define \tilde{u} to be the approximation in \mathcal{B} that minimizes the RRMSE over all such approximations in that basis. Specifically, define \mathcal{F} to be the space of possible approximations built from linear combinations of elements in \mathcal{B} , and then

$$\tilde{u} := \operatorname{argmin}_{\hat{u} \in \mathcal{F}} \operatorname{RRMSE}(\hat{u}). \quad (26)$$

Let \hat{u} be the approximation computed at the same iteration of BASE-PC in the basis \mathcal{B} using N samples to form a design matrix \mathbf{D} . Using these definitions we may show a useful result that flows through the restricted isometry constant (RIC), [21, 52], which is denoted here by $\rho_s(\mathbf{D})$ and is defined to be the smallest number satisfying

$$(1 - \rho_s(\mathbf{D}))\|\mathbf{c}\|_2^2 \leq \|\mathbf{D}\mathbf{c}\|_2^2 \leq (1 + \rho_s(\mathbf{D}))\|\mathbf{c}\|_2^2, \quad (27)$$

for all \mathbf{c} having at most s non-zero entries. Here, $\rho_s(\mathbf{D})$ yields a uniform bound on the spectral radius of the submatrices of \mathbf{D} formed by selecting any s columns. We occasionally shorten $\rho_s(\mathbf{D})$ to ρ_s , which should not be confusing in context. Related to an RIC is a restricted isometry property (RIP) that occurs when the RIC reaches a small enough threshold, and a RIP guarantees that ℓ_1 -minimization provides a stable approximation. An example of such a restricted isometry property is given in Theorem 1 from [48], restated here in our notation. This theorem shows that if $\rho_{2s} < 3/(4 + \sqrt{6})$, where s is a sparsity parameter corresponding to how many basis functions are useful in building a surrogate approximation, then a stable recovery is assured.

Theorem 1. [48] *Let $\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{B}|}$ represent the solution that produce \tilde{u} . Let $\mathbf{c}^{(s)}$ denote the best approximation to $\tilde{\mathbf{c}}$ in terms of minimizing $\|\tilde{\mathbf{c}} - \mathbf{c}^{(s)}\|_2$, where $\mathbf{c}^{(s)}$ has at most s non-zero entries. Let $\hat{\mathbf{c}}$ be the solution to (7), and let δ used to compute that solution, be chosen such that $\|\mathbf{W}(\mathbf{u} - \Psi\hat{\mathbf{c}})\| \leq \delta\|\mathbf{W}\mathbf{u}\|_2$. If*

$$\rho_{2s}(\mathbf{D}) < \rho_\star := 3/(4 + \sqrt{6}) \approx 0.4652,$$

then,

$$\begin{aligned} \|\tilde{\mathbf{c}} - \hat{\mathbf{c}}\|_2 &\leq \frac{c_1}{\sqrt{s}}\|\mathbf{c}^{(s)} - \tilde{\mathbf{c}}\|_1 + c_2\text{RMSE}(\tilde{u}); \\ \|\tilde{\mathbf{c}} - \hat{\mathbf{c}}\|_1 &\leq c_3\|\mathbf{c}^{(s)} - \tilde{\mathbf{c}}\|_1 + c_4\text{RMSE}(\tilde{u})\sqrt{s}, \end{aligned}$$

where c_1, c_2, c_3 , and c_4 depend only on ρ_{2s} .

We note that in the non-sparse case, we take $s \geq 0.5|\mathcal{B}|$ and the requirements on ρ_{2s} are less stringent. For example by Theorem 1 of [53] we could take $\rho_{2s} \leq 4/(6 + \sqrt{6}) \approx 0.4734$. We also note that in this case $2s \geq |\mathcal{B}|$, and so the condition here translates to requiring $\rho_{|\mathcal{B}|} < \rho_\star$, which is an isometry condition with no “restriction” to vectors of a particular sparsity. Utilizing

a RIC with $s = |\mathcal{B}|$ is useful here where we do not want to assume sparsity and still want to guarantee a stable solution to (7). We note that such a condition is also useful for guaranteeing solutions computed via least-squares regression [19, 11], although we do not consider such solutions here. We conclude this section by noting that the condition on δ is not generally an issue, as cross-validation is chosen so as to minimize $\text{RRMSE}(\hat{u})$, and when cross-validation has accurate validation, this loosely corresponds to minimizing $\|\tilde{\mathbf{c}} - \hat{\mathbf{c}}\|_2$, so that even if δ does not satisfy the condition, the bound on $\|\tilde{\mathbf{c}} - \hat{\mathbf{c}}\|_2$ will still be satisfied regardless of which δ is chosen.

4.1.4. Coupling

We assume that the aggregate samples at each iteration of correction sampling closely resemble an independent sample. Heuristically, this is justified as the introduced dependence is given in terms of (14), which is mild. Rigorously, we assume the existence of at least one of several couplings [54] between samples, one corresponding to that of the BASE-PC iterative correction samplings, and the other a set of independent samples drawn from a distribution, that considering (14) should closely coincide with the coherence-optimal distribution for the particular working basis at that iteration. Unfortunately, comparing dependent distributions and coupled independent distributions is difficult to interpret and analyze, and a method for constructing a coupling is currently unavailable. As a result, we assume that a desired coupling exists with a few parameters, leaving as an open problem the verification of the existence of such couplings, as well as any construction of such a coupling. We operate under the heuristic that our coupling is such that the coupled independent distribution is near the coherence-optimal distribution, which is validated by the correction sampling implied by (14).

Specifically, a coupling here refers to a joint distribution from which random variables are drawn, so that they are dependent in a way that is favorable. Here, we want random variables drawn via the correction sampling distributions to behave similarly to random variables drawn independently from a particular distribution, which for the moment we denote g_\star . As the coupled samples are drawn independently, we can deploy powerful existing analysis. As we can bound the error for solutions computed using the samples drawn from g_\star , we can in turn bound convergence for those drawn via the correction sampling. We note that coupling may be done between individual realizations of $\boldsymbol{\xi}^{(i)}$, or by coupling the entire pool of realized samples $\{\boldsymbol{\xi}^{(i)}\}_{i=1}^{N_k}$, as long as the coupled samples respect that they are drawn inde-

pendently from some g_\star . This provides significant freedom in how couplings may be identified or constructed.

We now present the couplings that we consider here. Let \mathbf{D} and \mathbf{D}_\star be design matrices associated with a common basis \mathcal{B} . Let \mathbf{D}_\star be generated from independent, identically distributed, random sampling, and $\mu_\star(s)$ be the coherence associated with this distribution and basis, as by (24). If there exists a β and $\kappa_t > 0$ such that

$$\mathbb{P}\left(\|\mathbf{D}^T \mathbf{D} - \mathbf{D}_\star^T \mathbf{D}_\star\| > t\right) \leq \beta \exp\left(-\kappa_t N_k \mu_\star^{-1}(|\mathcal{B}|) \log^{-1}(|\mathcal{B}|)\right), \quad (28)$$

then we say that \mathbf{D} is *non-sparse-coupled* to \mathbf{D}_\star with coupling constants β and κ_t . This coupling is so named as it is most useful when considering non-sparse recovery. Let subscript \mathcal{S} denote taking the submatrix associated with columns in \mathcal{S} . Another form of coupling is given by,

$$\sup_{|\mathcal{S}| \leq s} \mathbb{P}\left(\|\mathbf{D}_\mathcal{S}^T \mathbf{D}_\mathcal{S} - \mathbf{D}_{\mathcal{S},\star}^T \mathbf{D}_{\mathcal{S},\star}\| > t\right) \leq \beta \exp\left(-\kappa_t N_k \mu_\star^{-1}(s) \log^{-1}(|\mathcal{B}|) \log^{-3}(s)\right), \quad (29)$$

and if this holds, then we say that \mathbf{D} is *s-coupled* to \mathbf{D}_\star . This form of coupling is useful for considering recovery uniformly over coefficient supports in the case of sparse recovery. We also consider another form of coupling that is weaker than *s-coupling*, in that it requires the supremum of (29) to hold over a smaller set. Specifically, fix a set \mathcal{S}_0 , corresponding to a fixed support set that is good for building an approximation to the QoI. Define \mathcal{S}_r to be the set of $\mathcal{S} := \mathcal{S}_0 \cup \mathcal{R}$, where $|\mathcal{R}| \leq r$. Let \mathbf{D} , \mathbf{D}_\star , and μ_\star be as before. If there exists a β such that for some $\kappa_t > 0$,

$$\sup_{\mathcal{S} \in \mathcal{S}_r} \mathbb{P}\left(\|\mathbf{D}_\mathcal{S}^T \mathbf{D}_\mathcal{S} - \mathbf{D}_{\mathcal{S},\star}^T \mathbf{D}_{\mathcal{S},\star}\| > t\right) \leq \beta \exp\left(-\kappa_t N_k \mu_\star^{-1}(s+r) \log^{-1}(|\mathcal{B}|)\right), \quad (30)$$

then we say that \mathbf{D} is *(s, r)-coupled* to \mathbf{D}_\star . This recovery is useful for the non-uniform version of sparse recovery, that is, when we consider the recovery of a single QoI. As the set \mathcal{S}_r has comparatively fewer sets over which to take the supremum; the *(s, r)-coupling* is generally weaker than the *(s + r)-coupling*.

We remark again that the authors are unaware of how to identify such couplings or in how to bound the relevant β and κ parameters associated

with them. Intuitively, we expect the proposed sampling to behave similarly to independent sampling, and this framework can make the concept of similarity to independence explicit. Here, the difference between the iteratively adjusted sample and independent samples is by the relationship in (14), and so we expect the samples to behave similarly to independent samples, which is seen experimentally, where the two sets are indistinguishable in appearance.

The following theorem utilizes each of the above couplings to achieve a corresponding conclusion. Specifically, it links the non-independent random sampling that we use with the independent sampling that is a common assumption in most recovery theorems. This performs the heavy lifting for showing the recovery results in Sections 4.2 and 4.3. We note that while μ_\star , β_\star and κ'_t do depend on the couplings, and hence on k , we suppress this dependence for notational brevity.

Theorem 2. *For the k th iteration of sample expansion and solution computation, let \mathcal{B}_k denote the basis; N_k denote the total number of samples; and \mathbf{D}_k denote the design matrix. Fix $t > 0$, and assume that at least one of the three couplings (28), (29) or (30) exists, with the corresponding coupling constants for $t \geq \epsilon_t$ for some unspecified ϵ_t that is bounded away from zero. Let $s, |\mathcal{B}| > 1$. There exists $\kappa'_t > 0$ depending on t , β , and κ_t ; but independent of the other variables such that if non-sparse-coupling holds,*

$$\mathbb{P}\left(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{I}\| > t\right) \leq 2\beta_\star \exp\left(-\kappa'_t N_k \mu_\star^{-1}(|\mathcal{B}_k|) \log^{-1}(|\mathcal{B}_k|)\right), \quad (31)$$

where $\beta_\star = \max(\beta, 1)$. Let $\mathbf{D}_{k,\mathcal{S}}$ corresponds to the columns of \mathbf{D}_k corresponding to basis functions in \mathcal{S} . If the s -coupling of (29) holds then with κ'_t having the same dependency as before,

$$\sup_{|\mathcal{S}| \leq s} \mathbb{P}\left(\|\mathbf{D}_{k,\mathcal{S}}^T \mathbf{D}_{k,\mathcal{S}} - \mathbf{I}\| > t\right) \leq 2\beta_\star \exp\left(-\kappa'_t N_k \mu_\star^{-1}(s) \log^{-1}(|\mathcal{B}_k|) \log^{-3}(s)\right), \quad (32)$$

where $\beta_\star = \max(\beta, C_t)$ for some unspecified universal C_t . If the (s, r) -coupling of (30) holds with $r = Cs$, where C is near unity but has a mild dependence on $(s, |\mathcal{B}_k|, N_k, \mu_\star(s))$, then with κ'_t having the same dependency as before,

$$\sup_{\mathcal{S} \in \mathcal{S}_r} \mathbb{P}\left(\|\mathbf{D}_{k,\mathcal{S}}^T \mathbf{D}_{k,\mathcal{S}} - \mathbf{I}\| > t\right) \leq 2\beta_\star \exp\left(-\kappa'_t N_k \mu_\star^{-1}(s) \log^{-1}(|\mathcal{B}_k|)\right), \quad (33)$$

where $\beta_\star = \max(\beta, C_t)$, for some unspecified universal C_t having a minor dependence on $(s, |\mathcal{B}_k|, N_k, t)$.

Proof. We first define \mathbf{D}_{k^\star} to be a design matrix made from N_k samples drawn independently from the coupled distribution for samples at the k th iteration, denoted g_{k^\star} , using the basis \mathcal{B}_k . We consider first the non-sparse-coupling. We apply Lemma 2 to this matrix to get that there exists $\kappa'' > 0$, which in this case is a modest universal constant, such that

$$\mathbb{P}\left(\|\mathbf{D}_{k^\star}^T \mathbf{D}_{k^\star} - \mathbf{I}\| > t\right) \leq |\mathcal{B}_k| \exp\left(-\kappa'' t N_k \mu_\star^{-1}(|\mathcal{B}_k|)\right),$$

and that this holds for all $t \geq t_\epsilon$ for some unspecified $t_\epsilon > 0$. We now consider the coupling between the original matrix \mathbf{D}_k and its coupled, independently sampled matrix, \mathbf{D}_{k^\star} . This depends on the type of coupling considered, and we consider first the non-sparse-coupling. For a fixed t' , there exists a $\kappa_{t'}$, such that

$$\mathbb{P}(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{D}_{k^\star}^T \mathbf{D}_{k^\star}\| > t') \leq \beta |\mathcal{B}_k| \exp(-\kappa_{t'} N_k \mu_\star^{-1}(|\mathcal{B}_k|)).$$

Now,

$$\begin{aligned} \mathbb{P}\left(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{I}\| > t\right) &\leq \mathbb{P}\left(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{D}_{k^\star}^T \mathbf{D}_{k^\star}\| + \|\mathbf{D}_{k^\star}^T \mathbf{D}_{k^\star} - \mathbf{I}\| > t\right), \\ &\leq \min_{\substack{t_1 + t_2 = t \\ t_1 \geq t_\epsilon}} |\mathcal{B}_k| \left(\beta \exp(-\kappa_{t_1} N_k \mu_\star^{-1}(|\mathcal{B}_k|)) + \exp(-\kappa'' t_2 N_k \mu_\star^{-1}(|\mathcal{B}_k|))\right). \end{aligned}$$

Recall that κ'' is a universal constant. For some $\kappa'_t > 0$, dependent on t , and κ_{t_1} as a function of t_1 for $t_1 \in [\epsilon_t, t]$,

$$\mathbb{P}\left(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{I}\| > t\right) \leq 2\beta_\star |\mathcal{B}_k| \exp\left(-\kappa'_t N_k \mu_\star^{-1}(|\mathcal{B}_k|)\right),$$

where $\beta_\star = \max(\beta, 1)$. We consider the transfer of $|\mathcal{B}_k|$ into the exponential as $\log(|\mathcal{B}_k|)$ and note that for $|\mathcal{B}_k| > 1$ this can be handled by changing the constant κ'_t . For the case $|\mathcal{B}_k| = 1$, there is no need to move $|\mathcal{B}_k|$ into the exponential. As a result, this shows (31), giving for a newly defined κ'_t and $|\mathcal{B}_k| > 1$

$$\mathbb{P}\left(\|\mathbf{D}_k^T \mathbf{D}_k - \mathbf{I}\| > t\right) \leq 2\beta_\star \exp\left(-\kappa'_t N_k \mu_\star^{-1}(|\mathcal{B}_k|) \log^{-1}(|\mathcal{B}_k|)\right).$$

To show (32), (33), we assume either appropriate coupling and let $\mathbf{D}_{k^*,\mathcal{S}}$ denote the submatrix of \mathbf{D}_{k^*} corresponding to restricting to the columns associated with basis functions in \mathcal{S} . We must address the bound as a supremum over choices of \mathcal{S} . In the case of (32) a similar argument as above leads to,

$$\begin{aligned} \sup_{|\mathcal{S}| \leq s} \mathbb{P} \left(\|\mathbf{D}_{k,\mathcal{S}}^T \mathbf{D}_{k,\mathcal{S}} - \mathbf{I}\| > t \right) &\leq \min_{\substack{t_1+t_2=t \\ t_1 \geq \epsilon_t}} \left\{ \sup_{|\mathcal{S}| \leq s} \mathbb{P} \left(\|\mathbf{D}_{k,\mathcal{S}}^T \mathbf{D}_{k,\mathcal{S}} - \mathbf{D}_{k^*,\mathcal{S}}^T \mathbf{D}_{k^*,\mathcal{S}}\| < t_1 \right) \cdots \right. \\ &\quad \left. + \sup_{|\mathcal{S}| \leq s} \mathbb{P} \left(\|\mathbf{D}_{k^*,\mathcal{S}}^T \mathbf{D}_{k^*,\mathcal{S}} - \mathbf{I}\| < t_2 \right) \right\}, \end{aligned}$$

where the difference between showing (32) and (33) is taking a supremum over \mathcal{S} belonging to different sets.

The first term on the right hand side is already accounted for by the definition of s -coupling, but the second term is a subtle term to bound. The analogy between the first and second terms of the right hand side also occurs with regards to (33) and (s, r) -coupling. The couplings are defined in such a way that bounds for the first and second term are compatible so that the bounds in (32) and (33) are closely connected with bounds associated with the independent samples from g_* , with corrections to the constant β_* and κ'_t that account for the coupling. Specifically for s -coupling we claim that for some κ''_t and C_t depending only on t that

$$\sup_{|\mathcal{S}| \leq s} \mathbb{P} \left(\|\mathbf{D}_{k^*,\mathcal{S}}^T \mathbf{D}_{k^*,\mathcal{S}} - \mathbf{I}\| < t \right) \leq C_t \exp \left(-\kappa''_t N_k \mu_*^{-1}(s) \log^{-1}(|\mathcal{B}_k|) \log^{-3}(s) \right).$$

And that the s -coupling insures then that with the potential changes in constants that (32) holds. A similar bound holds for (s, r) -coupling and (33) with the optimization over a different set. We now argue that both such bounds hold for the independently generated rows, showing the theorem.

Recall that the coupling is defined such that the matrix \mathbf{D}_{k^*} has independent rows. This allows tighter bounds on this quantity than the naïve union bound over all sets satisfying $|\mathcal{S}| \leq s$, which would introduce a pessimistic order in the bound. The specifics of these tighter bounds are detailed and not presented here, but we point the interested reader to Talagrand's majorizing measures [55, 56] as well as works of Rudelson and Vershynin [57, 58, 59, 51]. We also point to Section 8.6 of [60] for results that more directly translate

to our use. We note that these results are typically presented in terms of the coherence parameter in (23), but that the proofs translate to those of (24). We also note that these results often go further and show results in terms of the sample size needed for an effective recovery, so that the result posted here is an intermediate result. For the most directly applicable results with respect to (32) we point to Theorem 8.4 of [60] and the closely connected Proposition 7.1 of [48]. For results directly applicable to (33) we reference Section 2.3 of [20] and the associated proofs. ■

We add some remarks regarding Theorem 2. We note here that the relationship of β_* in terms of the maximum of β and another parameter, and similarly the potential decrease of κ_t to κ'_t is because our analysis compares to independent samples. In the case that $C_t \geq \beta$, and a similar relationship on the associated coefficients of exponential decay, then the recovery requires a similar number of samples to that of independent sampling. It is not clear in practice how small the corresponding β and κ_t values can be, but we note that if the sampling is itself independent, then the trivial coupling of samples to themselves yields a value of $\beta = 0$. In such a case, or in the relaxed case that β is below some threshold while κ_t is above some threshold, the recovery bound may be reworked to be identical to independent sampling, modulo a constant factor of 2 by the proof technique here. As a result, if the non-independent sample used in BASE-PC resembles an independent sampling, then it may be expected that the β and κ_t values are small enough that the independent sampling result dominates the recovery. This appears to hold for the examples in Section 3, and it is suspected to hold in some generality.

We additionally note that this theorem is written in terms of $\mu_*(s)$, corresponding to the definition of coherence in (24), this may be bounded in terms of the other coherence definitions via Lemma 1. We also note that these bounds are problematic if s or $|\mathcal{B}| = 1$, but that this is an artifact of bounds within the proof, and could be removed by replacing the corresponding $\log(1)$ terms with 1.

Theorem 2 is sufficient to bound errors for each iteration, and we are equipped to show results for recovery in both the sparse and non-sparse cases.

4.2. Sparse Recovery

Here we consider the recovery of solutions when the sparsity parameter s satisfies $s < 0.5|\mathcal{B}_k|$. We consider the case of uniform and non-uniform recovery, where uniform recovery refers to the ability of the matrix \mathbf{D}_k to

recover any signal of sparsity s . This first result corresponds to uniform recovery. We recall from Lemma 1 that $\mu_\star(s) \leq s\mu_\infty$ and $\mu_\star(s) \leq \mu_2$. The coherence-optimal relationship in practice leads to $\mu_\star(2s)$ being proportional or nearly proportional to $2s$, and the ℓ_2 -coherence optimal sampling used here insures that $\mu_\star(2s) \leq |\mathcal{B}_k|$ for all s .

Corollary 1. Uniform Sparse Recovery: *Let $t < 3/(4 + \sqrt{6})$, and assume that a $2s$ -coupling holds with regards to Theorem 2. For some C , let N_k be such that,*

$$N_k \geq (C + \log(2\beta_\star))(\kappa'_t)^{-1}\mu_\star(2s)\log^3(2s)\log(|\mathcal{B}_k|). \quad (34)$$

Then for the k th iteration of BASE-PC, it follows that, with probability

$$p_k \geq 1 - \exp(-C), \quad (35)$$

the computed surrogate \hat{u}_k satisfies

$$\text{RRMSE}(\hat{u}_k) \leq D_1 \text{RRMSE}(\tilde{u}_k) + D_2 \frac{\|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1}{\sqrt{s}\sqrt{\mathbb{E}(u^2(\boldsymbol{\xi}))}}, \quad (36)$$

where D_1, D_2 are constants that depend only on t ; \hat{u}_k is the approximation computed via BASE-PC; and \tilde{u}_k is an optimal approximation as in (26). This result holds uniformly over any $\tilde{\mathbf{c}}_k$.

Remark 1. *We note that $\|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1/\sqrt{\mathbb{E}(u^2(\boldsymbol{\xi}))}$ converges to zero as $s \rightarrow |\mathcal{B}_k|$, even without the \sqrt{s} term. Though we do not write it that way, this result may be used with a minimization over a range of s such that (34) is satisfied.*

Proof. As the assumptions of Theorem 2 is satisfied, we may rewrite (32) as

$$\begin{aligned} \log \left(\sup_{|S| \leq 2s} \mathbb{P} \left(\|\mathbf{D}_{k,S}^T \mathbf{D}_{k,S} - \mathbf{I}\| > t \right) \right) &\leq \log(2\beta_\star) - \kappa'_t N_k \mu_\star^{-1}(2s) \log^{-1}(|\mathcal{B}|) \log^{-3}(2s), \\ &\leq \log(2\beta_\star) - (C + \log(2\beta_\star)), \\ &= -C. \end{aligned}$$

where the second inequality follows from (34). From this it follows that

$$\begin{aligned} \sup_{|S| \leq 2s} \mathbb{P} \left(\| \mathbf{D}_{k,S}^T \mathbf{D}_{k,S} - \mathbf{I} \| > t \right) &\leq \exp(-C), \\ \sup_{|S| \leq 2s} \mathbb{P} \left(\| \mathbf{D}_{k,S}^T \mathbf{D}_{k,S} - \mathbf{I} \| \leq t \right) &\geq 1 - \exp(-C). \end{aligned}$$

We note that if $\sup_{|S| \leq 2s} \| \mathbf{D}^T \mathbf{D} - \mathbf{I} \| \leq t$ then it follows that $\rho_{2s} \leq t$. We may then apply Theorem 1 to get that,

$$\begin{aligned} \|\hat{\mathbf{c}} - \tilde{\mathbf{c}}\| &\leq \frac{c_3}{\sqrt{s}} \|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1 + c_2 \text{RMSE}(\tilde{u}_k); \\ \sqrt{\mathbb{E}(\hat{u}(\Xi) - \tilde{u}(\Xi))^2} &\leq \sigma_{\min}(\mathbf{D})^{-1} \|\hat{\mathbf{c}} - \tilde{\mathbf{c}}\|, \\ &\leq (1-t)^{-1} \left(\frac{c_3}{\sqrt{s}} \|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1 + c_2 \text{RMSE}(\tilde{u}_k) \right); \\ &\leq \frac{C_3}{\sqrt{s}} \|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1 + C_2 \text{RMSE}(\tilde{u}_k), \end{aligned}$$

where the precise value of C_2 depends on t , and c_2 , and similarly C_3 depends on c_3 and t . As

$$\begin{aligned} \sqrt{\mathbb{E}(\hat{u}(\Xi) - u(\Xi))^2} &\leq \sqrt{\mathbb{E}(\hat{u}(\Xi) - \tilde{u}(\Xi))^2} + \sqrt{\mathbb{E}(\tilde{u}(\Xi) - u(\Xi))^2}, \\ &\leq (C_2 + 1) \text{RMSE}(\tilde{u}_k) + \frac{C_3}{\sqrt{s}} \|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1. \end{aligned}$$

Dividing both sides by $\sqrt{\mathbb{E}(u^2(\Xi))}$ and setting $D_1 = (C_2 + 1)$, $D_2 = C_3$, shows (36). ■

We may also address non-uniform recovery which theoretically requires fewer samples, and is especially useful within the context of UQ where design matrices are rarely used to recover large numbers of vastly differing QoIs. We show this scaling in the next result, which shows that several log terms may be removed from N_k .

Corollary 2. Non-Uniform Sparse Recovery: *Assume that the (s, r) -coupling holds with regards to Theorem 2. Let $t = 1/4$, for some C let N_k be such that,*

$$N_k \geq (C + \log(2\beta_\star))(\kappa'_t)^{-1} \mu_\star(s + r) \log(|\mathcal{B}_k|). \quad (37)$$

Using the same notation as Theorem 1 and Corollary 1, we have for the k th iteration of BASE-PC, with probability

$$p_k \geq 1 - \exp(-C), \quad (38)$$

the computed surrogate \hat{u}_k satisfies

$$\text{RRMSE}(\hat{u}_k) \leq D_1 \text{RRMSE}(\tilde{u}_k) + D_2 \frac{\|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1}{\sqrt{s} \sqrt{\mathbb{E}(u^2(\boldsymbol{\xi}))}}, \quad (39)$$

where D_1, D_2 are constants that depend on t and have a mild dependence on $(s, |\mathcal{B}_k|, N_k)$; \hat{u}_k is the approximation computed via BASE-PC; and \tilde{u}_k is an optimal approximation as in (26).

Remark 2. We note that the use of $t = 1/4$ is for compatibility with the theory presented in [20], and that this value could be taken larger. We also note that the dependency of D_1 and D_2 on $(s, r, N, |\mathcal{B}_k|)$ is never larger than $D_3 \log^2(|\mathcal{B}_k|)$ for some unspecified D_3 . Finally, as in Corollary 1, a similar optimization over s may be performed.

Proof. This proof is similar to that of Corollary 1, but utilizing the fact that we are restricting our coefficient support. Technical details are omitted, as the proof relies on different optimizations and estimates that are generally more favorable with regards to the constants. We point the interested reader to [20] and the proofs leading up to Theorem 1.3 there, as those are sufficient. We note how that paper is presented mostly in terms of the LASSO estimator, which is a dual form of the ℓ_1 -minimization problem in (7), but the results translate without issue. We bring special attention to the weak RIP of Section 2.3 of that paper which is the motivation for the (s, r) -coupling and its use in Theorem 2. ■

Corollaries 1 and 2 suggest that a number of samples that scales nearly linearly with the sparsity of the problem is sufficient to guarantee recovery, and demonstrates requirements for stability with respect to the correction sampling described in Section 2.2.1. As seen in Section 3, the BASE-PC iteration often selects bases with a number of elements that scale nearly linearly with the number of samples, suggesting that in those cases the desired sparsity parameter is a fraction of the total number of basis functions.

4.3. Non-Sparse Recovery

Here we consider the recovery of solutions when the sparsity parameter s satisfies $s \geq 0.5|\mathcal{B}|$. We note that the results in this section use the non-sparse-coupling. We note here that the independent coherence-optimal sampling gives $\mu_2(|\mathcal{B}_k|) = |\mathcal{B}_k|$, which is the theoretical minimum. The correction sampling aims to admit a coupling so that $\mu_\star(|\mathcal{B}_k|)$ remains near $|\mathcal{B}_k|$.

Corollary 3. Non-Sparse Recovery: *Let $t = 3/(4 + \sqrt{6})$, and assume that non-sparse-coupling holds and that the assumptions of Theorem 2 and Theorem 1. For some C , let N_k be such that,*

$$N_k \geq (C + \log(2\beta_\star))(\kappa'_t)^{-1}\mu_\star(|\mathcal{B}_k|)\log(|\mathcal{B}_k|). \quad (40)$$

Then for the k th iteration of BASE-PC it follows that, with probability

$$p_k \geq 1 - \exp(-C), \quad (41)$$

the computed surrogate \hat{u}_k satisfies

$$\text{RRMSE}(\hat{u}_k) \leq D_1 \text{RRMSE}(\tilde{u}_k), \quad (42)$$

where D_1 , depends only on t ; \hat{u}_k is the approximation computed via BASE-PC; and \tilde{u}_k is an optimal approximation as in (26). We note that this result holds uniformly over all \tilde{u}_k .

Proof. There are no significant differences between this proof and that of Corollary 1. The primary difference is that the proof of Corollary 1 requires a supremum over certain support sets, while this proof assumes the largest possible support, which leads to a bound that is more favorable than setting $s = |\mathcal{B}|$ in Corollary 1. We note that in the context of that Corollary, $\|\mathbf{c}_k^{(s)} - \tilde{\mathbf{c}}_k\|_1 = 0$, which explains the disappearance of the corresponding D_2 term. ■

4.4. Dimension Independent Scaling

We conclude this section with a theoretically satisfying guarantee for sufficiently smooth u , specifically that recovery may be achieved with a number of samples that does not depend on the dimension of the problem. If an exponentially decaying bound can be guaranteed for the coefficients, then to reach a particular RRMSE, $|\mathcal{B}|$ scales independently of dimensionality,

and so too does the necessary N to achieve a particular RRMSE, assuming the basis is approximately identified, and that a quality non-sparse-coupling exists.

Lemma 4. *Let \mathbf{i} be the $d \times 1$ vector that indexes the order of the basis function in each of d dimensions, and let $c_{\mathbf{i}} = \mathbb{E}(u(\Xi)\psi_{\mathbf{i}}(\Xi))$ denote the corresponding coefficient for the most accurate reconstruction of the surrogate, \hat{u} . If there exists $B > 0$ and $\alpha > 0$ such that*

$$|c_{\mathbf{i}}| \leq B \exp \left(-\alpha \sum_{k=1}^d k^2 i_k \right), \quad (43)$$

then for any $\epsilon \in (0, 0.9)$, there exists an anisotropic order basis, \mathcal{B}_{ϵ} , such that

$$|\mathcal{B}_{\epsilon}| \leq \epsilon^{-\nu}, \quad (44)$$

where $\nu \geq 0$ depends only on D , α and $\mathbb{E}(u^2(\Xi))$. With $\tilde{u}_{\mathcal{B}}$ as in (26), it follows that

$$\text{RRMSE}(\tilde{u}_{\mathcal{B}}) \leq \epsilon. \quad (45)$$

Remark 3. *We note that requiring $\epsilon \in (0, 0.9)$ is due to an estimate involving $\log(\epsilon)$ that may produce issues for ϵ near 1. Specifically, in (44), for ϵ near 1 we would expect a basis having 1 basis function to suffice. While this could be guaranteed with an arbitrary basis, this is difficult to guarantee with the anisotropic order basis, as the only available such basis is the basis with $\mathbf{p} = \mathbf{0}$, which is a basis consisting only of a constant term. In some cases, this basis function may not contribute to an accurate approximation, that is \tilde{u} built in this basis may still have $\text{RRMSE}(\tilde{u}) = 1$. Bounding ϵ away from 1 removes this issue, which is not of much practical interest when compared to the case of ϵ approaching 0. Another fix to the issue would be a bound such as $C\epsilon^{-\nu}$, but we avoid this approach due to an already large number of constants being used in this analysis.*

Proof. Consider using exact projection coefficients, i.e $c_k = \mathbb{E}(u(\Xi)\psi_k(\Xi))$, and including the basis functions associated with largest magnitude coefficients until the RRMSE is less than ϵ . We bound the size of such a desired basis by considering how many terms of the sum, $k i_k$, return values less than

any given threshold M , that is we consider all basis functions associated with \mathbf{i} satisfying

$$\sum_{k=1}^d k^2 i_k \leq M. \quad (46)$$

We note that this set of basis functions corresponds to an anisotropic order basis with each $p_k = M/k^2$. We may bound the number of such functions, denoted by B_M , independently of dimension. Specifically, the set of all terms that have non-zero order in exactly one dimension is bounded by,

$$M \sum_{k=1}^{\infty} k^{-2} = \frac{M\pi^2}{6},$$

which follows from (46) by considering how many i_k satisfy the relationship in each dimension, with

$$\sum_{k=1}^{\infty} k^{-2} = \pi^2/6,$$

being a classical result. Combinatorially, we may then bound the set of terms with non-zero order in exactly l dimensions that satisfy (46) by $(M\pi/6)^l/l!$, that is for \mathbf{i}_l having at most l non-zero entries,

$$\left| \left\{ \mathbf{i}_l : \sum_{k=1}^d k^2 i_k = M \right\} \right| \leq \left(\frac{M\pi^2}{6} \right)^l (l!)^{-1}. \quad (47)$$

Note that the constant term in the basis, $\mathbf{i} = \mathbf{0}$, corresponds to including indices with non-zero order in zero dimensions. Summing over basis functions that include elements in any of l dimensions for $l \geq 0$ gives that,

$$B_M \leq \sum_{l=0}^{\infty} \left(\frac{M\pi^2}{6} \right)^l (l!)^{-1} = \exp \left(\frac{M\pi^2}{6} \right), \quad (48)$$

which we note does not depend on d .

We consider now how large M should be to insure that $\text{RRMSE}(\tilde{u}_{\mathcal{B}})$ is below ϵ . From (47) we can define \tilde{u}_M to be the function approximation that uses all coefficients satisfying (46), and note that with (43),

$$\text{MSE}(\tilde{u}_M) \leq B^2 \sum_{l>M} \left(\frac{\pi^2 e^{-\alpha}}{6} \right)^{2l} (l!)^{-2}.$$

The convergence here is spectral as $M \rightarrow \infty$. Thus there exists a $\nu > 0$, depending on α and B , such that for any $\epsilon \in (0, 0.9)$, and for all $M \geq -\nu \log(\epsilon)$,

$$\text{MSE}(\tilde{u}_M) \leq \epsilon.$$

We note that to strengthen this to the RRMSE in (45), we may still take $M \geq -\nu \log(\epsilon)$, and need only potentially increase ν while adding a dependence on $\mathbb{E}(u^2(\Xi))$.

These two results bound M sufficiently for (45), and the number of basis functions that satisfy (46). Hence we have that \mathcal{B}_ϵ satisfying $\text{RRMSE}(\tilde{u}_B)$ is an anisotropic order basis defined by taking each

$$p_k = M/k^2 = -\nu \log(\epsilon)/k^2.$$

Hence, using $M = -\nu \log(\epsilon)$ in (48) it follows that for any $\epsilon \in (0, 0.9)$, that $|\mathcal{B}_\epsilon| \leq \epsilon^{-\nu}$ where ν depends on α , B , and $\mathbb{E}(u^2(\Xi))$. We remark that these p_k are not necessarily integers, and that requiring p_k to be integers would in turn require a modest increase to ν . ■

The following corollary then shows that the influence of dimensionality has the potential to be significantly reduced when considering basis adaptation. Under the assumptions of Lemma 4, an anisotropic order basis exists for an accurate approximation with a number of basis functions independent of dimension, which implies that a number of samples to guarantee an accurate computation is also independent of dimension. Note that computations in Section 2 scale favorably in dimension due to the d parameters to define the anisotropic total order basis, so that the computations in the BASE-PC iteration scale well with dimension.

This identifies a class of problems where BASE-PC may achieve accurate results with a benign scaling in dimension. The issue that prevents a stronger statement to this effect is that there is no guarantee provided that such a basis can be identified by the BASE-PC iteration. However, the search of BASE-PC that continually minimizes the estimate of $\text{RRMSE}(\hat{u})$ is reasonable, and in practice it has consistently found quality bases.

Corollary 4. *Let the assumptions of Lemma 4 and Theorem 2 be satisfied such that there exists a non-sparse-coupling at each iteration of BASE-PC. Let $\epsilon \in (0, 0.9)$. Let $t < 3/(4 + \sqrt{6})$. There exists a $\nu > 0$, and an anisotropic*

order basis \mathcal{B}_ϵ satisfying (44). Fix $\epsilon \in (0, 0.9)$. If at the k th iteration of BASE-PC it follows that $\mathcal{B}_\epsilon \subset \mathcal{B}_k$, and \hat{u} is an approximation in \mathcal{B}_k computed using

$$N_k \geq -\nu \log(\epsilon)(C + \log(2\beta_\star))(\kappa'_t)^{-1} \mu_\star(|\mathcal{B}_k|). \quad (49)$$

samples then it follows that, with probability

$$p_k \geq 1 - \exp(-C), \quad (50)$$

the computed surrogate, \hat{u}_k satisfies

$$\text{RRMSE}(\hat{u}) \leq \epsilon. \quad (51)$$

Remark 4. We note that in the event a coupling exists such that β_\star , κ'_t and $\mu_\star(|\mathcal{B}_k|)$ are independent of d the dimension of the Ξ , then no statement of this corollary depends on d . That is, if the problem exhibits a certain decay in the importance of dimension and order, made explicit in Lemma 4; and quality couplings exist to independent samples, as from Theorem 2; then there exists an anisotropic order basis such that it is possible to guarantee recovery for problems of arbitrarily high dimensions with a finite number of basis functions and samples.

Proof. From Lemma 4, we have a bound on the size of the desired basis as it scales with ϵ for an anisotropic order basis, \mathcal{B}_ϵ given as in Lemma 4, and satisfying (44), so that

$$|\mathcal{B}| \leq \epsilon^{-\nu}; \quad \log(|\mathcal{B}|) \leq -\nu \log(\epsilon).$$

To insure (51) holds, take a larger basis for \mathcal{B}_ϵ whose optimal approximation has an error of ϵ/D_1 , where D_1 is as from Corollary 3. This effect guarantees that (51) holds while requiring a further increase in ν , due to the need for a larger basis. Note that κ'_t is as from Theorem 2, when using the non-sparse-coupling. Then with this basis, and a number of samples satisfying (49), the computed approximation satisfies (51) with probability at least as large as in (50). ■

5. Conclusions

A definition for anisotropic order [6] basis is presented as being compatible with accurate PC expansions and having a number of parameters that scales

as the problem dimension, allowing a tunable basis which limits the number of unnecessary basis functions in our active basis while still admitting accurate approximations. Using this basis, an adaptive-sampling is identified so that at each iteration all samples taken up to that point are effectively used in the computation of the surrogate solution. If the basis adaptation is successful, then we have performed a theoretical analysis for both sparse and non-sparse recovery. Further, under some assumptions, when recovering a solution from a class of smooth functions, both the size of a necessary basis and the overall number of samples necessary to compute surrogates to desired accuracy does not depend on the dimension of the problem, representing a significant result with respect to the so-called curse of dimensionality.

Also, as the design matrix has fewer basis functions and samples than standard PCE approaches, the computation of the coefficients needed to construct the surrogate scales relatively well with the dimension of the problem. Although no guarantee is provided that a successful basis adaptation can be identified in any given number of basis adaptation iterations, the deployed heuristic of greedily searching to minimize an estimate of $\text{RRMSE}(\hat{u})$ is numerically seen to perform well for the examples considered. The scaling is significantly more favorable than the exponential growth in basis functions when considering total order expansions.

Numerically we see that a smoothness of the problem in terms of its polynomial coefficients is more informative of the success of this method than the dimensionality of the problem, and that for problems which utilize high-order basis functions that the proposed correction sampling is of significant assistance for recovering a quality approximation when compared to sampling from the orthogonality distribution.

Acknowledgements

The work of JH was supported by the DARPA EQuiPS project.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advances Scientific Computing Research, under Award Number DE-SC0006402, and NSF grant CMMI-145460.

References

- [1] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer Verlag, 1991.

- [2] O. L. Maitre, O. Knio, Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics, Springer, 2010.
- [3] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.
- [4] D. Xiu, G. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644.
- [5] C. Soize, R. Ghanem, Physical systems with random uncertainties: Chaos representations with arbitrary probability measure, SIAM Journal of Scientific Computing 26 (2) (2005) 395–410.
- [6] J. Bäck, F. Nobile, L. Tamellini, R. Tempone, Spectral and High Order Methods for Partial Differential Equations: Selected papers from the ICOSAHOM '09 conference, June 22-26, Trondheim, Norway, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, Ch. Stochastic Spectral Galerkin and Collocation Methods for PDEs with Random Coefficients: A Numerical Comparison, pp. 43–62.
- [7] F. Ni, P. Nguyen, J. F. G. Cobben, Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow, IEEE Transactions on Power Systems PP (99) (2016) 1–1. [doi:10.1109/TPWRS.2016.2558622](https://doi.org/10.1109/TPWRS.2016.2558622).
- [8] A. Doostan, H. Owhadi, A. Lashgari, G. Iaccarino, Non-adapted sparse approximation of PDEs with stochastic inputs, Tech. Rep. Annual Research Brief, Center for Turbulence Research, Stanford University (2009).
- [9] A. Doostan, H. Owhadi, A non-adapted sparse approximation of PDEs with stochastic inputs, Journal of Computational Physics 230 (2011) 3015–3034.
- [10] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, Journal of Computational Physics 280 (2015) 363–386.

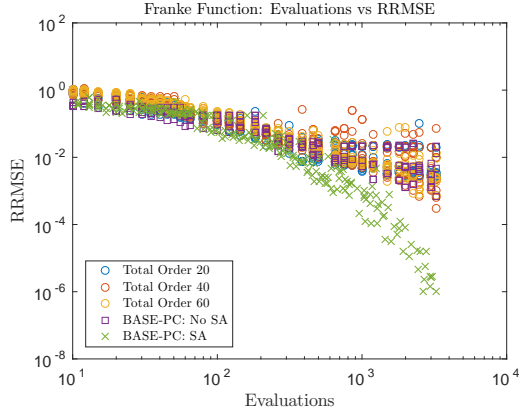
- [11] J. Hampton, A. Doostan, Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression, *Computer Methods in Applied Mechanics and Engineering* 290 (2015) 73–97.
- [12] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1998) 33–61.
- [13] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* 43 (1) (2001) 129–159.
- [14] D. Donoho, Compressed sensing, *IEEE Transactions on information theory* 52 (4) (2006) 1289–1306.
- [15] A. Bruckstein, D. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Review* 51 (1) (2009) 34–81.
- [16] E. van den Berg, M. P. Friedlander, Probing the pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing* 31 (2) (2008) 890–912.
- [17] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing: Closing the gap between performance and complexity, *Tech. rep., DTIC Document* (2008).
- [18] X. Wan, G. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *J. Comp. Phys.* 209 (2005) 617–642.
- [19] A. Cohen, M. A. Davenport, D. Leviatan, On the stability and accuracy of least squares approximations., *Foundations of Computational Mathematics* 13 (5) (2013) 819–834.
- [20] E. J. Candès, Y. Plan, A probabilistic and riplless theory of compressed sensing, *Information Theory, IEEE Transactions on* 57 (11) (2010) 7235–7254.
- [21] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *Information Theory, IEEE Transactions on* 52 (2) (2006) 489–509.

- [22] M. Elad, Sparse and redundant representations: from theory to applications in signal and image processing, Springer, 2010.
- [23] Y. Eldar, G. Kutyniok, Compressed sensing: theory and applications, Cambridge University Press, 2012.
- [24] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics* 230 (2011) 2345–2367.
- [25] L. Mathelin, K. Gallivan, A compressed sensing approach for partial differential equations with random input data, *Commun. Comput. Phys.* 12 (2012) 919–954.
- [26] L. Yan, L. Guo, D. Xiu, Stochastic collocation algorithms using ℓ_1 -minimization, *International Journal for Uncertainty Quantification* 2 (3).
- [27] X. Yang, G. E. Karniadakis, Reweighted ℓ_1 minimization method for stochastic elliptic differential equations, *Journal of Computational Physics* 248 (2013) 87–108.
- [28] G. Karagiannis, G. Lin, Selection of polynomial chaos bases via bayesian model uncertainty methods with applications to sparse approximation of pdes with stochastic inputs, *Journal of Computational Physics* 259 (2014) 114–134.
- [29] J. Peng, J. Hampton, A. Doostan, A weighted ℓ_1 -minimization approach for sparse polynomial chaos expansions, *Journal of Computational Physics* 267 (2014) 92–111.
- [30] D. Schiavazzi, A. Doostan, G. Iaccarino, Sparse multiresolution regression for uncertainty propagation, *International Journal for Uncertainty Quantification* 4 (4) (2014) 303–331.
- [31] K. Sargsyan, C. Safta, H. Najm, B. Debusschere, D. Ricciuto, P. Thornton, Dimensionality reduction for complex models via bayesian compressive sensing, *International Journal for Uncertainty Quantification* 4 (2013) 63–93.

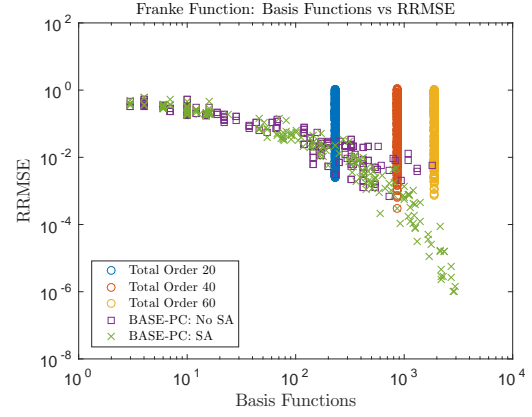
- [32] J. Jakeman, M. Eldred, K. Sargsyan, Enhancing ℓ_1 -minimization estimates of polynomial chaos expansions using basis selection, *Journal of Computational Physics* 289 (0) (2015) 18 – 34.
- [33] B. Adcock, Infinite-dimensional ℓ_1 minimization and function approximation from pointwise data, *arXiv preprint arXiv:1503.02352*.
- [34] H. Rauhut, R. Ward, Interpolation via weighted ℓ_1 minimization, *Applied and Computational Harmonic Analysis* 40 (2) (2016) 321–351.
- [35] J. Peng, J. Hampton, A. Doostan, On polynomial chaos expansion via gradient-enhanced ℓ_1 -minimization, *Journal of Computational Physics* 310 (2016) 440 – 458.
- [36] B. Adcock, A. C. Hansen, Generalized sampling and infinite-dimensional compressed sensing, *Foundations of Computational Mathematics* (2015) 1–61.
- [37] O. Le Maitre, H. Najm, R. Ghanem, O. Knio, Multi-resolution analysis of Wiener-type uncertainty propagation schemes, *J. Comp. Phys.* 197 (2) (2004) 502–531.
- [38] X. Wan, G. E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, *SIAM Journal on Scientific Computing* 28 (3) (2006) 901–928.
- [39] F. Ni, P. Nguyen, J. F. G. Cobben, Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow, *IEEE Transactions on Power Systems* PP (99) (2016) 1–1.
- [40] N. Alemazkour, H. Meidani, Divide and conquer: an incremental sparsity promoting compressive sampling approach for polynomial chaos expansions, *arXiv preprint arXiv:1606.06611*.
- [41] O. Dykstra, The augmentation of experimental data to maximize $[x \ x]$, *Technometrics* 13 (3) (1971) 682–688.
- [42] Y. Shin, D. Xiu, On a near optimal sampling strategy for least squares polynomial regression, *Journal of Computational Physics* 326 (2016) 931 – 946.

- [43] R. Franke, A critical comparison of some methods for interpolation of scattered data, Tech. rep., DTIC Document (1979).
- [44] O. LeMaitre, M. Reagan, H. Najm, R. Ghanem, O. Knio, A stochastic projection method for fluid flow. ii: Random process, *J. Comp. Phys.* 181 (2002) 9–44.
- [45] P. L. Quéré, Accurate solutions to the square thermally driven cavity at high rayleigh number, *Computers & Fluids* 20 (1) (1991) 29–41.
- [46] A. Makeev, D. Maroudas, I. Kevrekidis, Coarse stability and bifurcation analysis using stochastic simulators: Kinetic Monte Carlo examples, *The Journal of chemical physics* 116 (23) (2002) 10083–10091.
- [47] O. L. Maître, H. Najm, R. Ghanem, O. Knio, Multi-resolution analysis of wiener-type uncertainty propagation schemes, *J. Comput. Phys* 197 (2004) 502–531.
- [48] H. Rauhut, R. Ward, Sparse legendre expansions via ℓ_1 -minimization, *J. Approx. Theory* 164 (5) (2012) 517–533.
- [49] A. Narayan, J. D. Jakeman, T. Zhou, A christoffel function weighted least squares algorithm for collocation approximations, arXiv preprint arXiv:1412.4305.
- [50] J. A. Tropp, User-friendly tail bounds for sums of random matrices, *Foundations of Computational Mathematics* 12 (4) (2012) 389–434.
- [51] R. Vershynin, Compressed sensing: theory and applications, Ch. Introduction to the non-asymptotic analysis of random matrices, in: [23].
- [52] E. J. Candès, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathématique* 346 (9) (2008) 589–592.
- [53] S. Foucart, A note on guaranteed sparse recovery via ℓ_1 -minimization, *Applied and Computational Harmonic Analysis* 29 (1) (2010) 97–103.
- [54] T. Lindvall, Lectures on the coupling method, Courier Corporation, 2002.

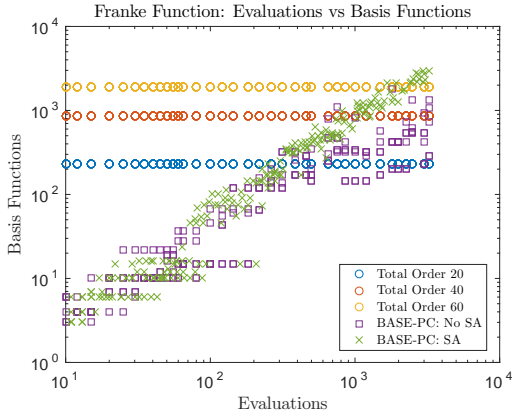
- [55] M. Talagrand, Majorizing measures: the generic chaining, *The Annals of Probability* (1996) 1049–1103.
- [56] M. Talagrand, Majorizing measures without measures, *Annals of probability* (2001) 411–417.
- [57] M. Rudelson, Almost orthogonal submatrices of an orthogonal matrix, *Israel Journal of Mathematics* 111 (1) (1999) 143–155.
- [58] M. Rudelson, Random vectors in the isotropic position, *Journal of Functional Analysis* 164 (1) (1999) 60–72.
- [59] M. Rudelson, R. Vershynin, Sampling from large matrices: An approach through geometric functional analysis, *Journal of the ACM (JACM)* 54 (4) (2007) 21.
- [60] H. Rauhut, Compressive sensing and structured random matrices, *Theoretical foundations and numerical methods for sparse recovery* 9 (2010) 1–92.



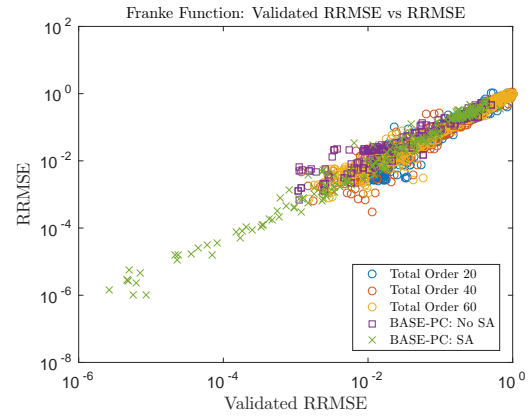
(a) RRMSE vs QoI evaluations



(b) RRMSE vs number of basis elements

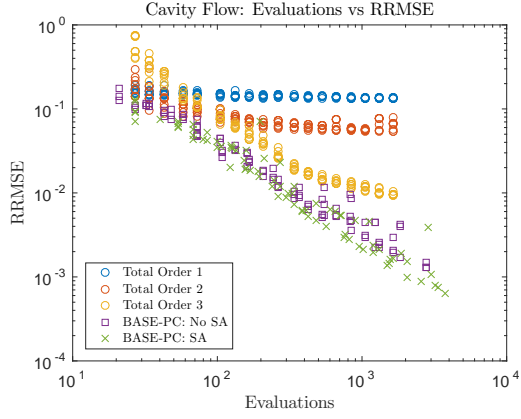


(c) QoI evaluations vs number of basis elements

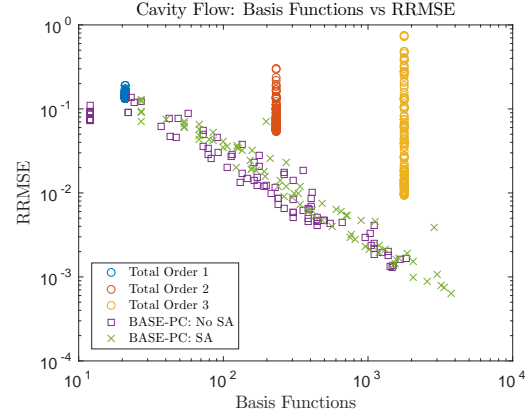


(d) RRMSE vs Estimated RRMSE

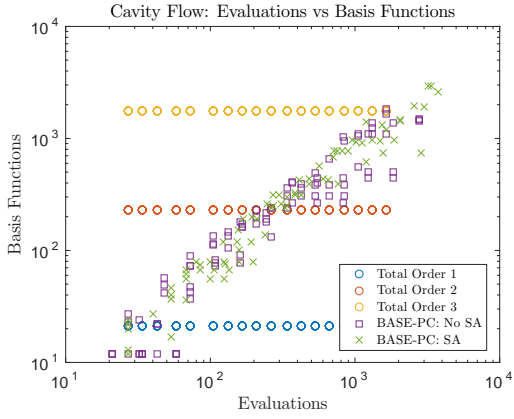
Figure 2: Comparisons of different methods for the Franke function.



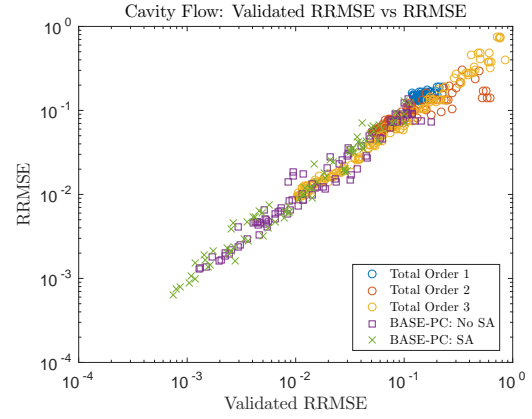
(a) RRMSE vs QoI evaluations



(b) RRMSE vs number of basis elements

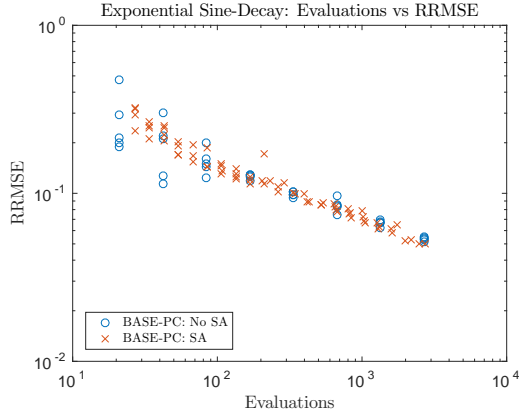


(c) QoI evaluations vs number of basis elements

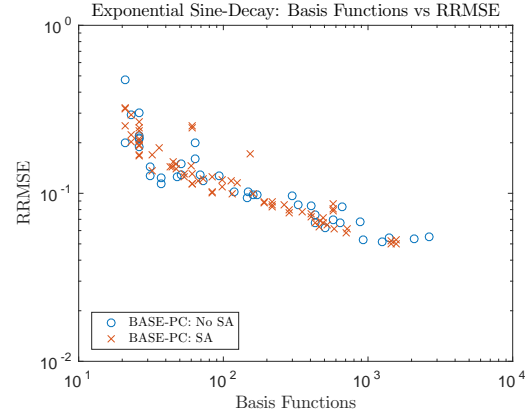


(d) RRMSE vs Estimated RRMSE

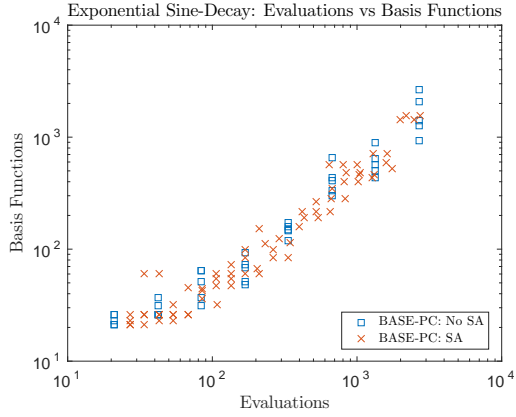
Figure 4: Comparisons of different methods for a cavity flow model with $d = 20$.



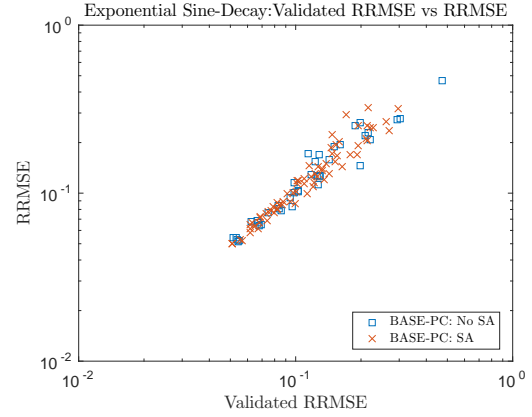
(a) RRMSE vs QoI evaluations



(b) RRMSE vs number of basis elements

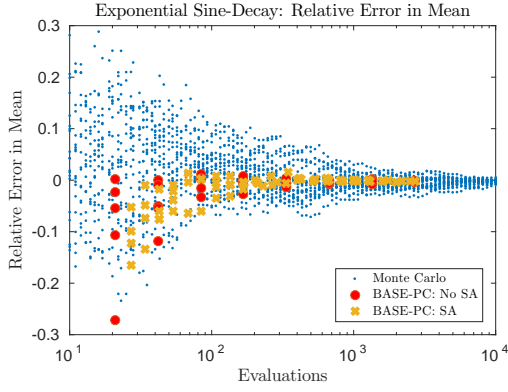


(c) QoI evaluations vs number of basis elements

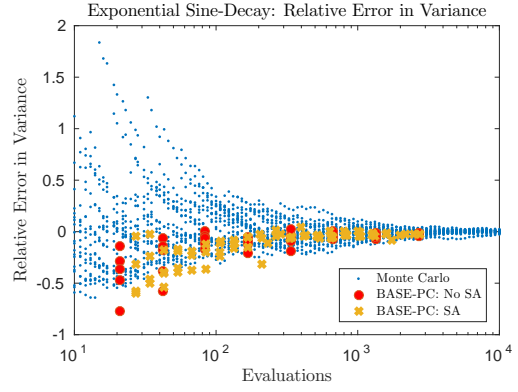


(d) RRMSE vs Estimated RRMSE

Figure 5: Comparisons of different methods for (20) with $d = 1000$.



(a) Mean vs QoI evaluations



(b) Variance vs QoI evaluations

Figure 6: Estimates for mean and variance for (20) with $d = 1000$.

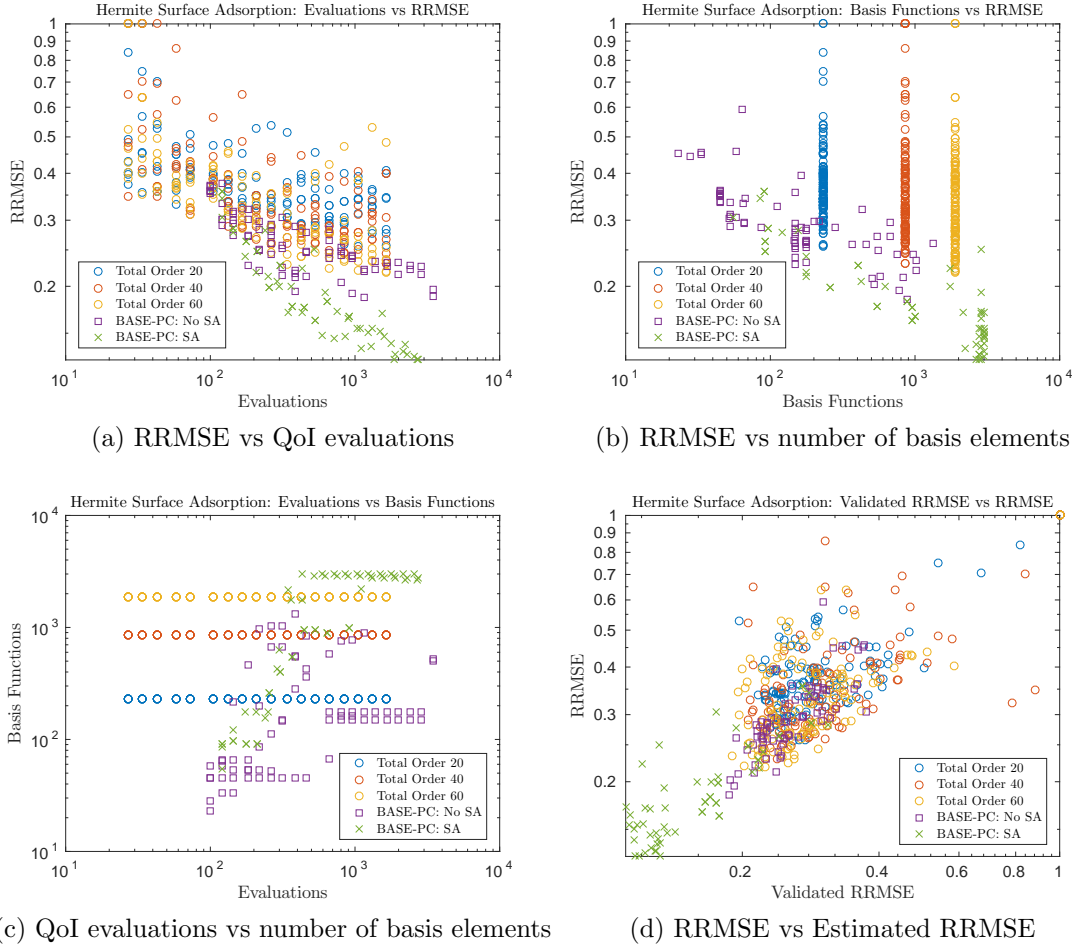
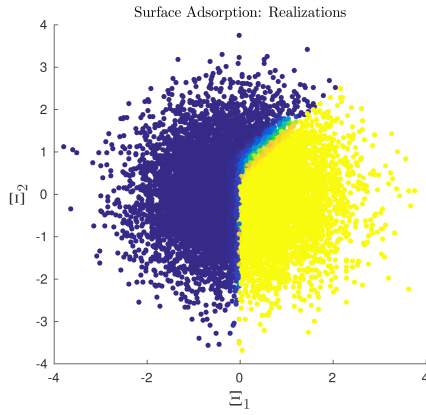
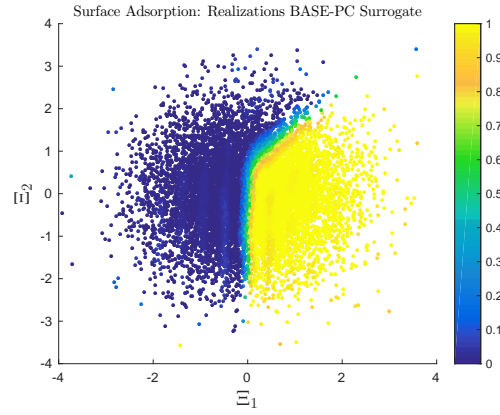


Figure 7: Comparisons of different methods for the surface adsorption model with Hermite polynomials.



(a) QoI from surface adsorption model



(b) Sample adaptive BASE-PC surrogate

Figure 8: Comparison of QoI with sample adaptive BASE-PC surrogate